The disc embedding theorem

Based on work of Michael H. Freedman

Editors: 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

Preface

In 1982, Michael Freedman, building upon 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 twelve 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* organised 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 https://www.mpim-bonn.mpg.de/node/4436.

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 organised 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 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 Programme in Topology: Christopher Davis, Peter Feller, Duncan McCoy, Jeffrey Meier, Allison 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 synthesised 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 computerised 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 of 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**, **Bižs94**, **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 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 organisation

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, Slava Krushkal, Andy Putman, and András Stipsicz.

Contents

List of figures

CHAPTER 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 to such questions is a *Poincaré complex*, roughly speaking a finite cell complex that satisfies n-dimensional Poincaré duality.

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*:

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

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 $\pi_1(X)$ 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 also 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.

Summarising, 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 realised by closed, smooth (equivalently, PL [**HM74**; **FQ90**, Theorem 8.3B]) or topological 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 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 $\mathbb{CP}^3$. 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 realised 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 non-trivial intersections as well, but we ignore those for now. If the hyperbolic pair could be represented by framed, embedded spheres which intersect exactly once,

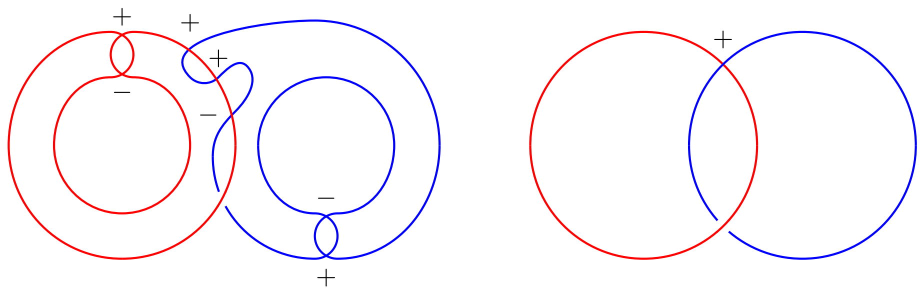


Figure 1.1. Trying to surger a hyperbolic pair. Left: Immersed spheres, depicted schematically, which intersect each other algebraically once but geometrically thrice. Right: The desired situation, where we have embedded spheres which intersect geometrically once.

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 to carrying out this strategy in dimension four lies in geometrically realising the algebraic intersection number, passing, as it were, from the left to the right of Figure 1.1. In the smooth category, Donaldson's diagonalisation 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 stabilisation 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, after possibly stabilisation by adding cancelling 2- and 3-handle pairs, that these 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 into 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 [**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 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.

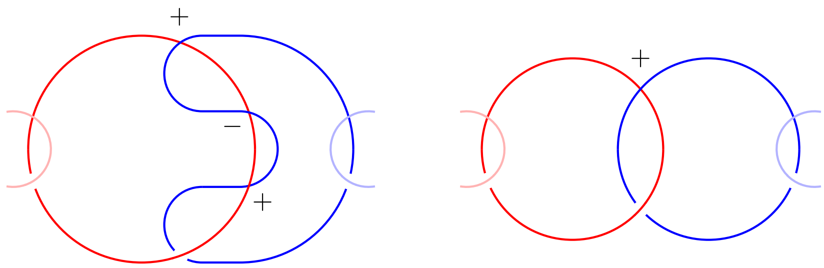


Figure 1.2. An algebraically dual pair consisting of a 2-handle belt sphere (red) and a 3-handle attaching sphere (blue) are shown. The light curves denote the corresponding geometrically dual spheres. Left: The belt sphere and the attaching sphere intersect algebraically once but geometrically thrice.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

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 realise 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.

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 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 $d \geq 3$,

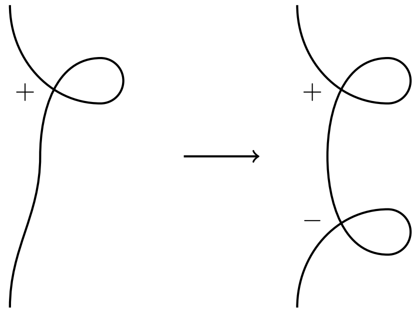


Figure 1.3. Adjusting the algebraic self-intersection number of an immersed submanifold by adding local kinks.

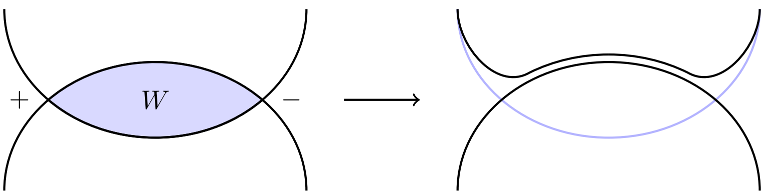


Figure 1.4. The Whitney move. Left: A Whitney disc W is shown in blue. Right: The Whitney move across W removes two intersection points.

$\pi_1(\mathrm{Gr}_{d-1}(\mathbb{R}^{2d-2})) \cong \mathbb{Z}/2$, 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) 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 ideal situations such as when the signs of the intersection points do not cancel or we have the wrong framing, i.e. one that does not extend over a Whitney disc.

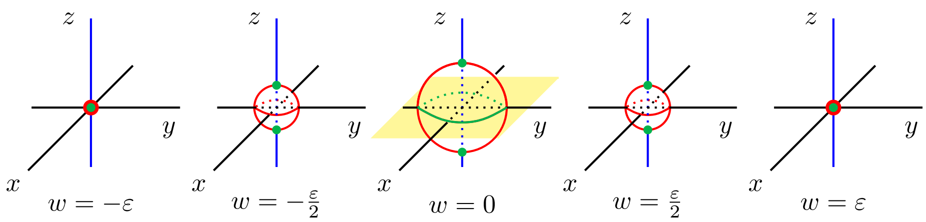


Figure 1.5. The Hopf link at a transverse intersection. Each of the five images above shows the $\mathbb{R}^3$-slice of $\mathbb{R}^4$ corresponding to the w-coordinate as indicated. Only the central image, where $w = 0$, contains the xy-plane, shown in yellow. The vertical lines in blue trace out the 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 when $w = \pm\varepsilon$. 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.

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

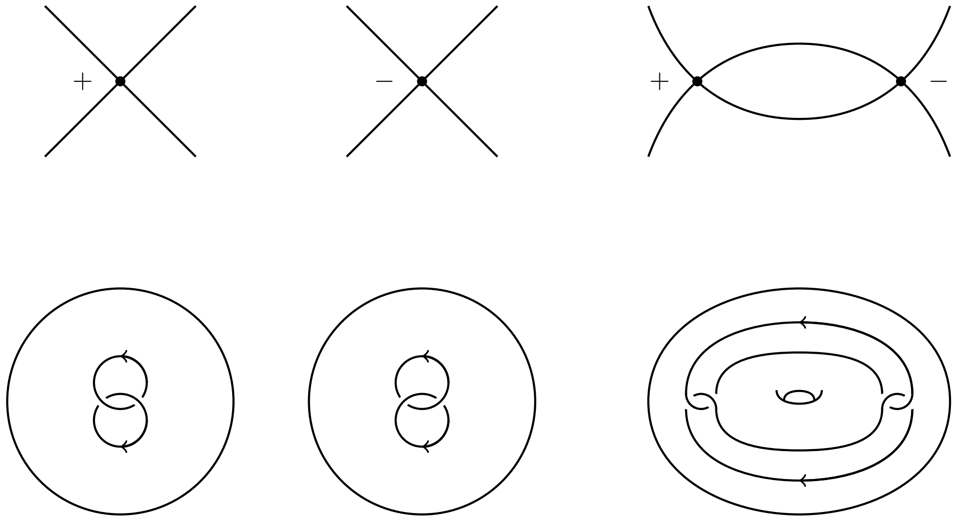


Figure 1.6. Banding together, with untwisted framing, two Hopf links lying in disjoint copies of S^3, around two intersection points produces a Bing double in $S^1 \times D^2 \subseteq S^1 \times S^2$. Each of the three lower pictures corresponds to the picture directly above it.

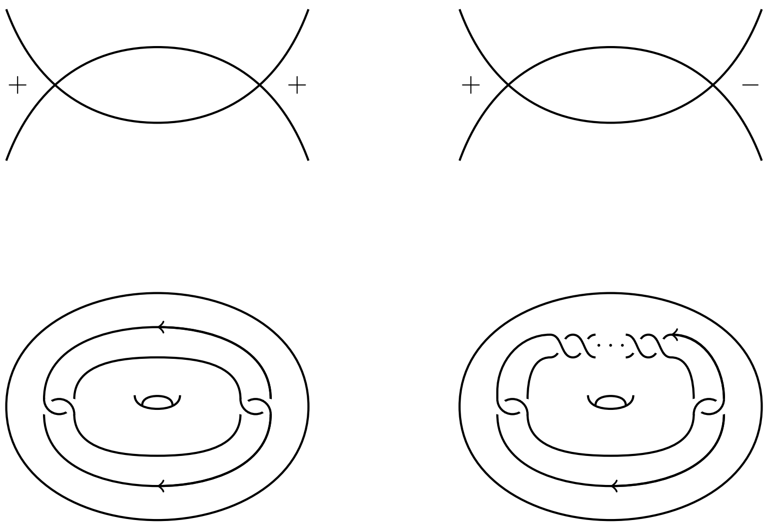


Figure 1.7. Left: When the paired intersection points have the same sign, we get a 2-component link with linking number ± 2. Right: An incorrectly framed Whitney circle produces a twisted Bing double of the Whitney circle.

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, 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**] realised 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 $\pi_1(M \smallsetminus A) \to \pi_1(M)$ is an isomorphism. Note that this implies that any curve in $M \smallsetminus A$ that extends to a map of a disc in M also extends to a map of a disc in the complement $M \smallsetminus A$. 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 as indicated in Figure 1.8 by so-called *finger moves*, 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, 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 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 $M \smallsetminus \widehat{S}$. 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 $\widehat{S}$ by 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.

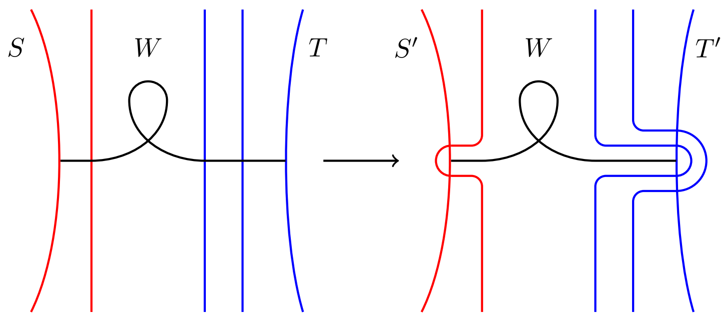


Figure 1.8. Trading intersections for self-intersections. A Whitney disc W (black) pairing algebraically cancelling intersection points between the spheres S (red) and T (blue) is shown in cross section. For each intersection of S with the interior of W, perform a homotopy of S to move the intersection 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.

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 perform 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.

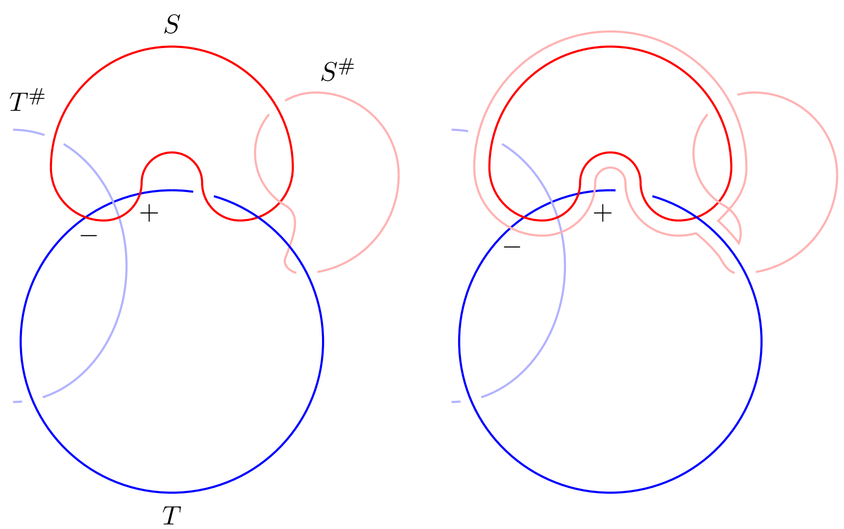


Figure 1.9. Adjusting the intersection number of $S^\#$ (light red) and T (blue), by tubing $S^\#$ into S (red) along a suitable arc. Do this for every point of intersection between $S^\#$ and T. Afterwards, the intersections of $S^\#$ and T are algebraically cancelling. These new intersections are marked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

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.

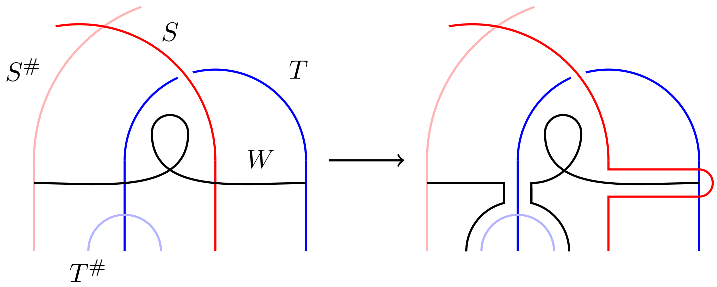


Figure 1.10. Obtaining a Whitney disc for intersections between $S^{\#}$ and T with interior in the complement of $S \cup T$. We see a Whitney disc W (black) pairing intersection points between $S^{\#}$ (light red) and T (blue). Remove intersection points of the interior of W with T by tubing W into 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.

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 (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 focussed 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 and embedded, and had the right framing, we could introduce an algebraically cancelling double point 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 by 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, where the boundaries of each stage other than the attaching region are removed. This is called a *Casson handle*. The attaching region of a Casson handle is the attaching region of any constituent Casson tower, which all coincide by definition. Thus, in the case of a simply connected ambient manifold, we have now replaced our immersed Whitney discs by 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 on to 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.

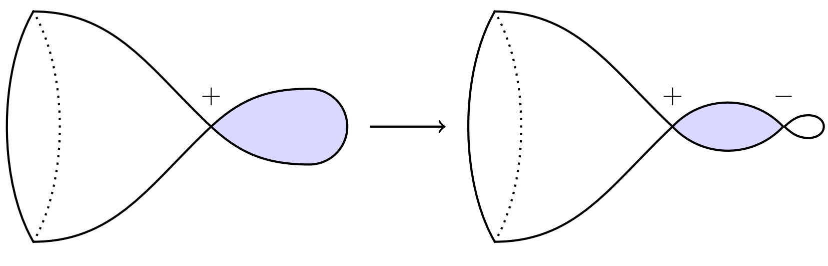


Figure 1.11. Whitney move to resolve a self-intersection. On the left we show a self-intersection point of a disc such that the double point loop bounds an embedded and framed disc whose interior is in the complement of the first disc. When we add a local kink of the opposite sign, a Whitney circle bounding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

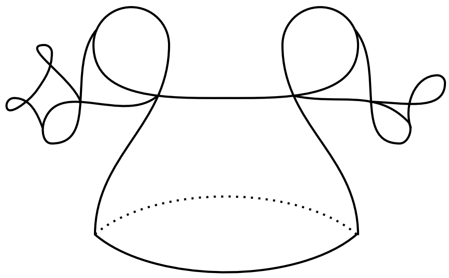


Figure 1.12. Schematic picture of the 2-dimensional spine of a Casson tower of height three.

Theorem 1.1 (Casson [**Cas86**, Lecture 1] (see also Siebenmann [**Sie80**])). *Every Casson handle is proper 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 as 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 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, the disc embedding theorem states that given a collection of immersed discs in a 4-manifold with algebraically dual spheres, nice enough fundamental group of the ambient manifold, and some restrictions on the intersections of the discs and the dual spheres, we may upgrade the immersed discs to embedded discs with the same boundary and tubular neighbourhoods, at the expense of leaving the smooth category.

The hypothesised 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 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 summarise 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 $\overline{g}_i$ 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 work of Kirby and Siebenmann [**KS77**, Essay 3, 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 realising 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 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, simply connected, topological 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, oriented, simply connected, topological* 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 realised 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 realised 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 realised as the intersection form of a closed, simply connected, topological 4-manifold.

It is still an open question exactly which indefinite forms are realised 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 realisation of even more bilinear forms as the intersection form 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 [**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] 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**, page 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

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

since $\operatorname{ks}(E_8) = 1$, $\operatorname{ks}(\mathbb{CP}^2) = \operatorname{ks}(\overline{\mathbb{CP}^2}) = 0$, 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, page 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 characterised 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 $\pi_1(S^4 \smallsetminus S) \cong \mathbb{Z}$, 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 mentioned earlier 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

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

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 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 $\mathbb{R}^4$. $\square$

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.

CHAPTER 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 his or her 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 $\{f_i\}$ by 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 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 visualise.

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.

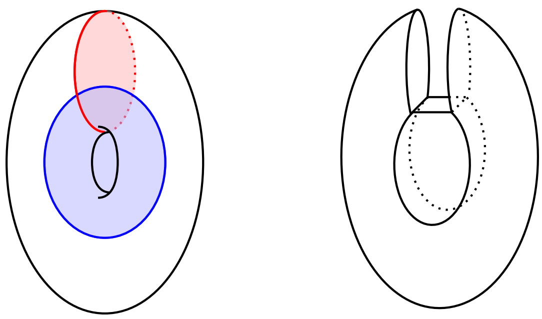
Step 1 (Modify the base discs until the intersections may be paired by Whitney discs, Proposition 16.1). We start with the initial hypothesised immersed discs $\{f_i\}$ and algebraically dual immersed spheres $\{g_i\}$, where $\{g_i\}$ has trivial intersection and self-intersection numbers by hypothesis. 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 $\{D_k'\}$ by 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 $M \smallsetminus \bigcup \nu f_i'$. 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 a neighbourhood of a transverse intersection 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, after contraction 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.



(a) (b)

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 summarise the current situation. We have replaced the hypothesised collection of immersed discs $\{f_i\}$ by 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 summarised in Figure 2.2.

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 work for a while with the sets $\{W_k\}$ and $\{\Sigma^c_k\}$.

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 $\{\Sigma^c_k\}$. 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 $\{\Sigma^c_k\}$, we can make the $\{W'_k\}$ and the caps of $\{\Sigma^c_k\}$ disjoint, so that the families $\{W'_k\}$ and $\{\Sigma^c_k\}$ 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 $\{\Sigma^c_k\}$ may change by a regular homotopy in this process, but we keep the same notation. The fact that the bodies of the $\{\Sigma^c_k\}$ lie in a neighbourhood of $\{f'_i\}$ plays an important rôle in the separation we have just performed. In particular, we needed the Whitney move to separate the caps of $\{W'_k\}$ and $\{\Sigma^c_k\}$, and the fact that the bodies of $\{\Sigma^c_k\}$ are located close to the boundary of the ambient manifold allowed us to assume that the relevant Whitney discs do not intersect these bodies. We postpone all further details to the actual proof.

Let us pause to describe an object called a *capped grope*. We could eliminate (in the fundamental group) the essential curves on any connected surface with a single boundary component by capping off a collection of curves that form a basis

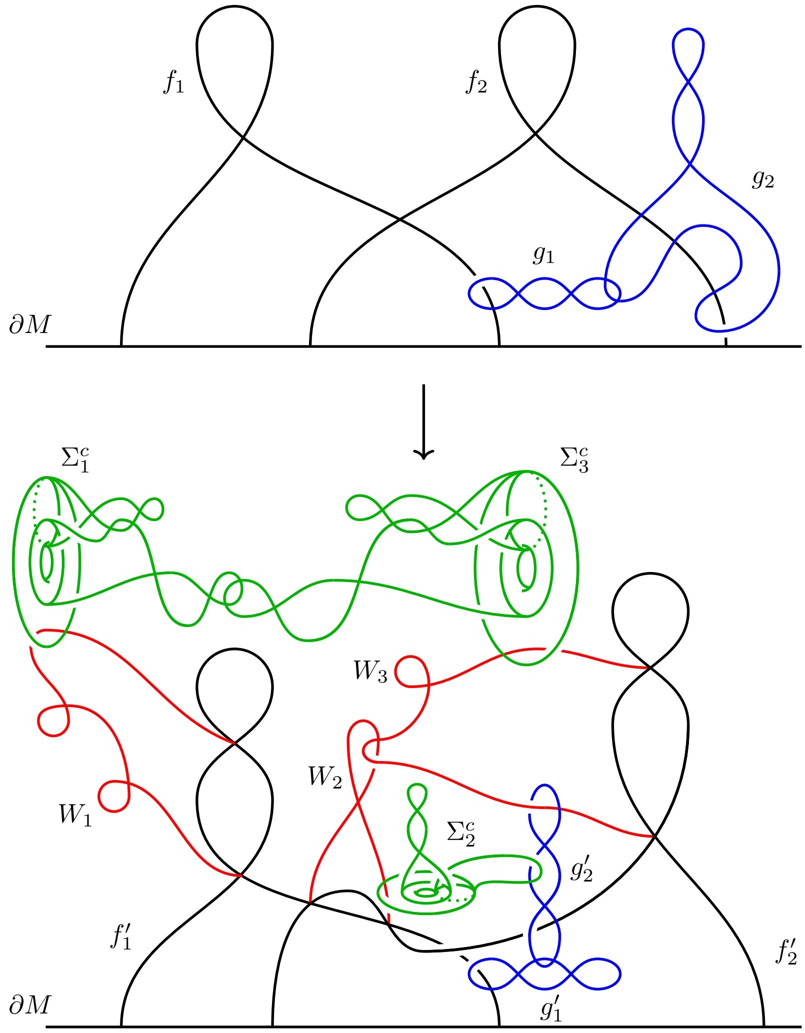


Figure 2.2. Summary of Proposition 16.1. The discs $\{f_i\}$ and $\{f_i'\}$ are in black, the spheres $\{g_i\}$ and $\{g_i'\}$ are in blue, the Whitney discs $\{W_k\}$ are in red, and the transverse capped surfaces $\{\Sigma_k^c\}$ are in green.

for the first homology with immersed discs. Alternatively, we could cap off such a collection of curves with further compact surfaces, so the fundamental group is generated by essential curves on this new layer of surfaces. Continue to add such surfaces iteratively to collections of curves forming a symplectic basis for the first homology of the previous layer of surfaces. Taking a 4-dimensional neighbourhood of the resulting object, at each stage, gives rise to a *grope*. A neighbourhood of the boundary of the base layer compact surface is called the *attaching region* of the grope. The fundamental group of a grope is generated by a collection of simple closed curves representing a basis for the first homology of the last layer of attached surfaces. Cap off these curves with plumbed (thickened) discs, to obtain a 4-dimensional object called a *capped grope*. The plumbed 2-handles attached at the latter step are called *caps*. Note that a capped grope with a single layer of surfaces is nothing more than a (neighbourhood of a) capped surface. The fundamental group of a capped grope is generated by the collection of double point loops of its caps. We describe gropes in much more detail in Section 12.1. Now we return to our outline of the proof.

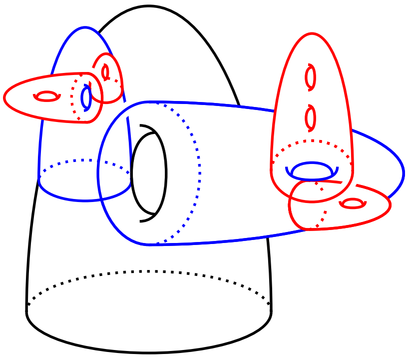


Figure 2.3. Schematic picture of the 2-dimensional spine of a height three grope.

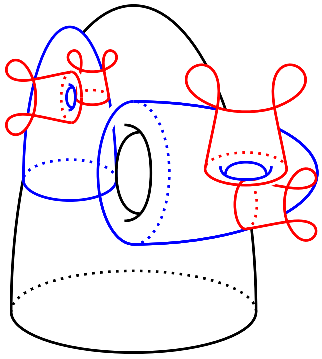


Figure 2.4. Schematic picture of the 2-dimensional spine of a height two capped grope.

Step 4 (Upgrade the Whitney discs further to capped gropes of height two, Proposition 16.2). By another round of tubing into the capped surfaces $\{\Sigma_k^c\}$, we transform $\{W_k'\}$ to capped gropes $\{G_k^c\}$ with two layers of surfaces. Note that gropes are 4-dimensional objects, so we take a neighbourhood of the surfaces and caps. Observe also that the caps of the gropes may intersect one another, but the gropes are otherwise disjoint. By creating and using another collection of geometrically transverse spheres, we once again ensure that the caps of $\{G_k^c\}$ do not intersect the caps of $\{\Sigma_k^c\}$, and as a result the collections are geometrically dual.

At this point, we have replaced the Whitney discs $\{W_k\}$ for the intersections and self-intersections of $\{f_i'\}$ by a collection of capped gropes $\{G_k^c\}$ with two layers of surfaces, equipped with a family of geometrically transverse capped surfaces $\{\Sigma_k^c\}$, all lying, apart from the attaching region of $\{G_k^c\}$, in the ambient manifold $M \smallsetminus \bigcup \nu f_i'$, which we know to have good fundamental group. Note that the bodies of the gropes are pairwise disjoint, but the caps may intersect. These two steps are performed in Proposition 16.2 and bring us in our summary to the end of Chapter 16.

STEP 5 (Upgrade the capped gropes to 1-storey capped towers, Proposition 17.1). In Proposition 17.3 we show that every capped grope with at least two layers of surfaces contains a new capped grope with arbitrarily many layers of compact surfaces and with the same attaching region. This procedure is called *Grope Height Raising*. We apply this procedure to the gropes $\{G^c_k\}$. The resulting tall gropes retain the set $\{\Sigma^c_k\}$ as geometrically dual capped surfaces. By the contraction and push off operations mentioned earlier, we ensure that the caps of the gropes are mutually disjoint and have algebraically cancelling double points. Observe that a capped grope has nontrivial fundamental group generated by the double point loops of its caps.

At this point of the proof, we use the fact that the ambient 4-manifold has good fundamental group to cap off these fundamental group generators with a new, second layer of plumbed thickened discs in the ambient 4-manifold. This is the only part of the proof that uses the good fundamental group hypothesis. A grope with two layers of plumbed discs attached at the top is called a 1*-storey capped tower*.

We create another round of geometrically dual spheres from parallel copies of the transverse capped surfaces $\{\Sigma^c_k\}$ and use them to ensure that the newest set of caps only intersect one another. This produces a collection of 1-storey capped towers $\{\mathcal{T}^c_k\}$, whose only intersections are among the last layer of caps, pairing the intersections and self-intersections of $\{f'_i\}$. Note, for future reference, that the constituent capped gropes of $\{\mathcal{T}^c_k\}$ may be assumed to have at least four surface stages, by Grope Height Raising.

It is finally time to bring the spheres $\{g'_i\}$, which recall are geometrically dual to the $\{f'_i\}$, out from storage. Right now, the collection $\{g'_i\}$ may intersect the 1-storey capped towers $\{\mathcal{T}^c_k\}$ arbitrarily. We contract the capped surfaces $\{\Sigma^c_k\}$ one last time to produce a final collection of geometrically transverse spheres $\{R_k\}$, this time for the 1-storey capped towers. Tube any intersections of the $\{g'_i\}$ with the 1-storey capped towers into the spheres $\{R_k\}$, to produce a new collection of immersed spheres $\{\overline{g}_i\}$. Note that the $\{f'_i\}$ and the $\{\overline{g}_i\}$ are geometrically dual, and moreover the 1-storey capped towers lie in the complement of both the $\{f'_i\}$ and the $\{\overline{g}_i\}$, except for where the attaching regions interact with $\{f'_i\}$. It is not too hard to see that the spheres $\{R_k\}$ are null-homotopic in M, since they arose from Clifford tori (Lemma 17.11), from which it follows that for each i, the sphere $\overline{g}_i$ is homotopic to g'_i, which is homotopic to g_i. The collection $\{\overline{g}_i\}$ is the one promised in the statement of the disc embedding theorem, and we will no longer modify this collection.

The construction so far is summarised in Proposition 17.12. We have now reached the end of the outline of Chapter 17.

From now on we completely forget the ambient 4-manifold and work solely within the 1-storey capped towers $\{\mathcal{T}^c_k\}$. Recall that the only intersections within the family $\{\mathcal{T}^c_k\}$ are amongst the caps. Our proof will be complete if we can show that every such collection of 1-storey capped towers contains pairwise disjoint 2-handles with the same attaching regions, and indeed this is what we shall do. First, we will show that any such collection of 1-storey capped towers contains a pairwise disjoint collection of skyscrapers with the same attaching regions.

Since the only intersections within the collection $\{\mathcal{T}^c_k\}$ are amongst the caps, the fundamental group of this collection is generated by double point loops in the caps. Recall that our guiding principle was to build a simply connected object, so we must somehow kill these curves. If we could replace the last layer of caps in a 1-storey capped tower by capped gropes, we would solve our current problem, but develop new essential double point loops at the caps of these newly added capped gropes. So we cap *those* off with new capped gropes, and so on. In other words, we iterate this process. If we stop after creating a finite number n of layers of capped gropes, the resulting object is called an *n-storey tower*, where each layer of capped gropes is called a *storey*. The *attaching region* of a tower is the attaching region of the base grope. We give a picture of the 2-dimensional spine of a 2-storey tower in Figure 2.5. As for capped gropes, the fundamental group of an n-storey capped tower is generated by double point loops at the caps of the last layer of capped gropes. Cap off these curves with plumbed thickened discs, to obtain a 4-dimensional object called an *n-storey capped tower*. The plumbed 2-handles attached at the last step are again called *caps*. The fundamental group of a capped tower is freely generated by a collection of double point loops of its caps.

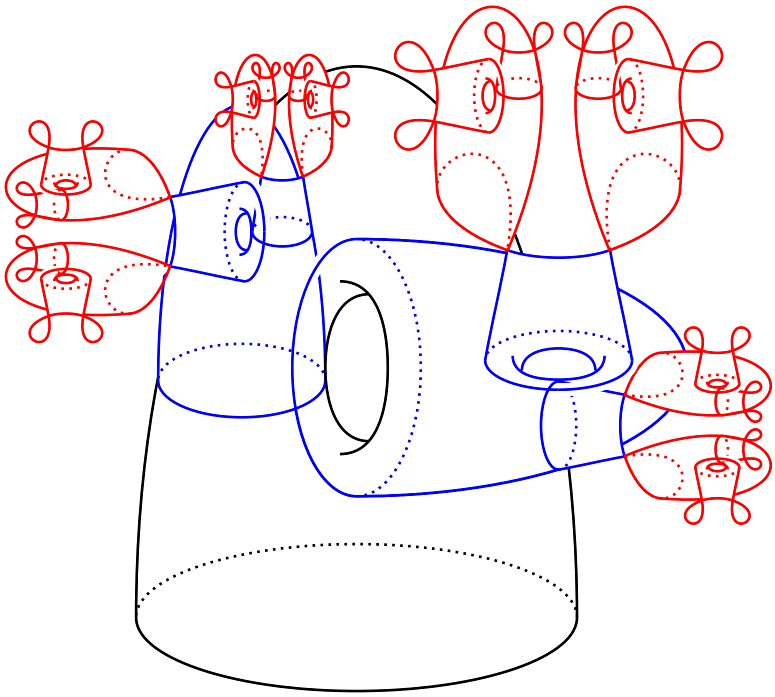


Figure 2.5. Schematic picture of the 2-dimensional spine of a 2-storey tower with two surface stages in the first storey and one surface stage in the second storey.

In Chapter 18, we show that we can find capped towers with arbitrarily many storeys and the same attaching regions within a collection of 1-storey capped towers with intersecting caps and at least four surface stages in each constituent capped grope. This is the *Tower Embedding Theorem Without Squeezing*. Note that if we had omitted the intervening surface stages in a tower we would have built a collection of Casson towers. However the surface stages are very useful, both for tower building and for other aspects of the argument, as we will see soon.

We have now seen how to build some very large iterated objects. However, while at each stage we eliminated some troublesome curves in the fundamental group, we keep finding new sets of problem curves. As for Casson towers, a key insight in this phase of the proof is that we do not need to stop at a finite number of stages in a tower. A tower with infinitely many layers of capped gropes is called an *infinite tower*, and is simply connected.

Let us address a second wish for our approximation of $D^2 \times D^2$, other than simple connectivity. While an infinite tower is simply connected, it is clearly noncompact and in particular, has infinitely many *ends*. Loosely speaking, a space has as many ends as essentially distinct rays homeomorphic to $[0, \infty)$ that leave every compact subset. For example the real line $\mathbb{R}$ has two ends while the plane $\mathbb{R}^2$ has a single end. The *endpoint compactification* of a space adds a point to each end and returns a compact space. For example, the endpoint compactification of $\mathbb{R}$ is homeomorphic to the closed interval $[0, 1]$, while that of $\mathbb{R}^2$ is homeomorphic to the sphere S^2. Ideally we would like to not only build an infinite tower but rather its endpoint compactification, called an *infinite compactified tower*. The endpoint compactification of a tower can be easily defined abstractly, but the difficult step is to find this compactification inside a given ambient manifold. For us, this will mean inside each 1-storey capped tower $\mathcal{T}_1^c$. In order to embed the endpoint compactification, we need some control over the higher storeys of a tower. We achieve this control in Chapter 18. We prove the Tower Embedding Theorem (Theorem 18.9), which shows that a collection of 1-storey capped towers with potentially intersecting caps and at least four surface stages in each constituent capped grope contains within it a collection of pairwise disjoint infinite compactified towers such that the higher storeys are contained in arbitrarily small balls. We also note here, for future application, that by contracting the top storey to discs, every 2-storey tower contains a 1-storey capped tower and thus contains an infinite compactified tower, with the same attaching region, such that the higher storeys are contained in arbitrarily small balls. By Grope Height Raising, we can also incorporate into the construction that each constituent capped grope has arbitrarily many layers of surfaces.

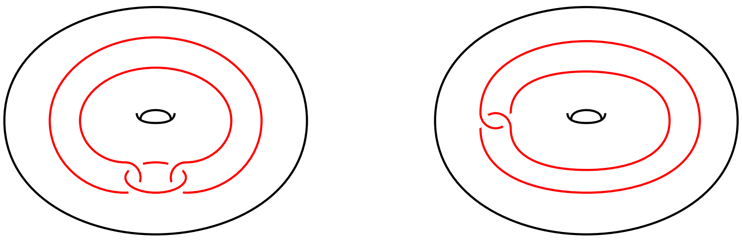
A *skyscraper* is an infinite compactified tower with two conditions on the number of surface stages in the constituent capped gropes. We require that there are at least four surface stages in each constituent capped grope, and that the number of surface layers in the tower should dominate the number of cap layers in a precise manner, which we do not state yet. In the next step of the proof, we replace the 1-storey capped towers found in the previous step by skyscrapers contained within them with the same attaching region.

Step 6 (Upgrade the 1-storey capped towers to skyscrapers, Theorem 18.9). Apply the Tower Embedding Theorem to the 1-storey capped towers obtained in the previous step of the proof. As mentioned earlier, this produces a pairwise disjoint collection of infinite compactified capped towers, with the same attaching regions, such that the higher storeys are contained in arbitrarily small balls. These infinite compactified towers must satisfy some further conditions on the number of surface stages to qualify as skyscrapers, but we can arrange for this by using Grope Height Raising.

To summarise, we have replaced the Whitney discs $\{W_k\}$ pairing the intersections and self-intersections of the $\{f_i'\}$ by pairwise disjoint skyscrapers whose interiors lie in the complement of the $\{f_i'\}$ and the $\{\overline{g}_i\}$. This brings us to the end of Chapter 18 and Part II. The proof so far is summarised in Proposition 18.12. Observe that to complete the proof of the disc embedding theorem it suffices to show that every skyscraper is homeomorphic to a standard 2-handle $D^2 \times D^2$ relative to its attaching region, since then the cores of these handles will provide framed disjointly embedded Whitney discs pairing the intersections and self-intersections of the $\{f_i'\}$. Performing the Whitney move over these will produce the desired collection of flat embedded discs $\{\overline{f}_i\}$ with the same framed boundary as the original discs $\{f_i\}$ equipped with the collection $\{\overline{g}_i\}$ of geometrically dual spheres. We already saw that the sphere g_i is homotopic to the sphere $\overline{g}_i$ for each i. The proof of the disc embedding theorem modulo the fact, proved in Part IV, that skyscrapers are standard, is given again at the end of Chapter 18.

2.3. Skyscrapers are standard

The next, and final, stage of the argument shows that every skyscraper is homeomorphic to $D^2 \times D^2$, relative to its attaching region. This takes place in Part IV. First we gather some facts about skyscrapers. Since skyscrapers are just a special type of infinite compactified tower, the Tower Embedding Theorem gives us the *Skyscraper Embedding Theorem* (Theorem 18.10): any two consecutive storeys of a skyscraper contains an embedded skyscraper with the same attaching region. Moreover, the boundary of a skyscraper is rather nice. We study skyscrapers and their boundaries using the techniques of *Kirby calculus* in Chapter 13. We will be particularly interested in the complement of the attaching region in the boundary of a skyscraper, which we call the *vertical boundary.* In Chapter 13, we learn that the vertical boundary for a tower with finitely many storeys is the complement of a link in a solid torus with many components. This link is obtained by performing successive iterations of ramified Bing and Whitehead doubling on the core of the solid torus (see Figures 2.6 and 2.7). Here, ramification corresponds to taking parallel copies of a link component. More precisely, we perform ramified Bing doubling for each layer of surfaces, and we perform ramified Whitehead doubling for each layer of discs within the tower. For an infinite tower, the vertical boundary is the complement in a solid torus of the infinite intersection of neighbourhoods of such doubled links.



(a) (b)

Figure 2.6. The Bing double (left) and a Whitehead double (right) of the core of a solid torus.

The ends of an infinite tower correspond to the number of branches in this construction, and so by definition the endpoint compactification corresponds to adding a single point per branch. One of the key inputs from Part I is that if the surface/Bing stages dominate the disc/Whitehead stages (corresponding to the second condition on the number of surface stages in a skyscraper), the infinite intersection above is such that, after endpoint compactification has added a point for each connected component, the resulting vertical boundary is homeomorphic to the solid torus again. The compactified vertical boundary can be identified with the quotient space of $D^2 \times S^1$, where each connected component of the infinite intersection is crushed to a point. So this quotient space of $D^2 \times S^1$ is homeomorphic to $D^2 \times S^1$. This is called *boundary shrinking* and is one of the reasons we insisted upon having surface stages in a skyscraper in the first place. For Casson towers/handles, the corresponding picture has no Bing steps, and we do not get boundary shrinking. Freedman's original proof circumvents this issue in a clever way, the key idea of

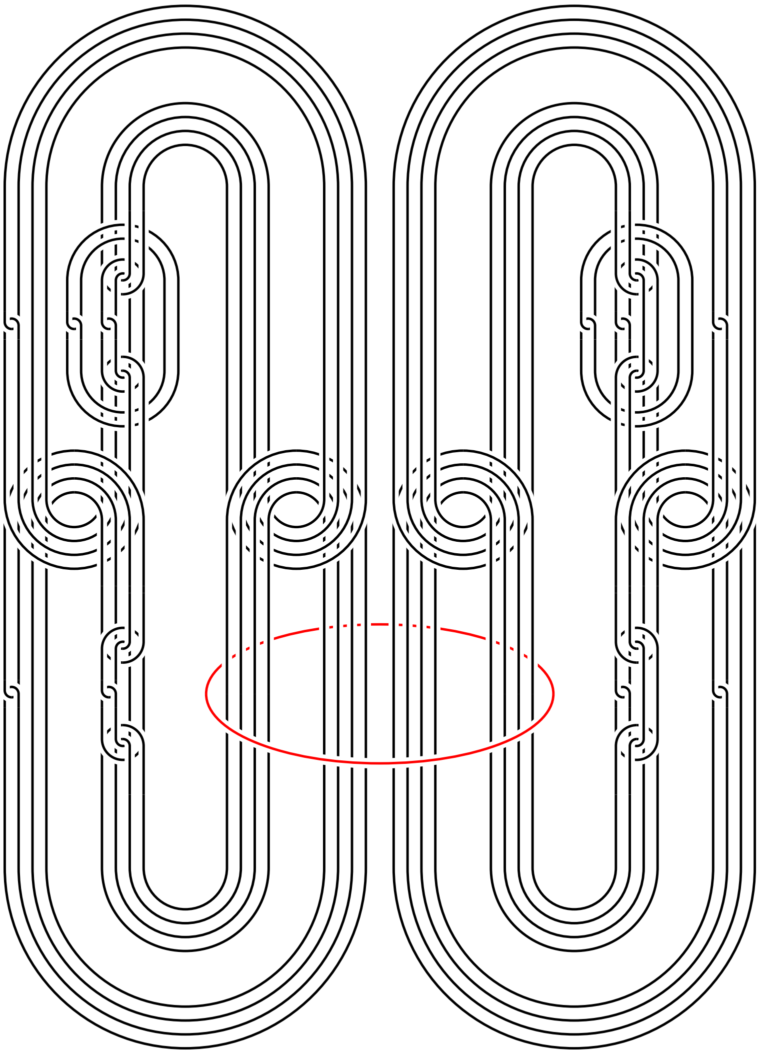


Figure 2.7. A link obtained by performing ramified Bing and Whitehead doubling on the core of a solid torus, namely the complement of the red circle in S^3.

which we describe in Chapter 7 (Chapter 7 is not necessary for our proof that skyscrapers are standard).

The practical upshot of the previous paragraph is that the vertical boundary of a skyscraper is homeomorphic to $D^2 \times S^1$. Moreover, from these last few facts it follows that the total boundary of a skyscraper is homeomorphic to S^3, which is a valuable sanity check that we are indeed on the right track towards showing that every skyscraper is homeomorphic to $D^2 \times D^2$.

Returning to the proof, recall once again that our only remaining task is to show that a skyscraper, denoted by $\widehat{\mathcal{S}}$, is homeomorphic to $D^2 \times D^2$, relative to the attaching region. The strategy is to find a sufficiently large common subset of both spaces, and for each space, quotient by identifying the closures of the connected components of the complement of the common subset to points. We will show that the two quotients are homeomorphic, and that the two original spaces $\widehat{\mathcal{S}}$ and $D^2 \times D^2$ are homeomorphic to their respective quotients. Since the proof is reasonably complicated, we give a more detailed sketch of the remaining steps in Chapter 27.

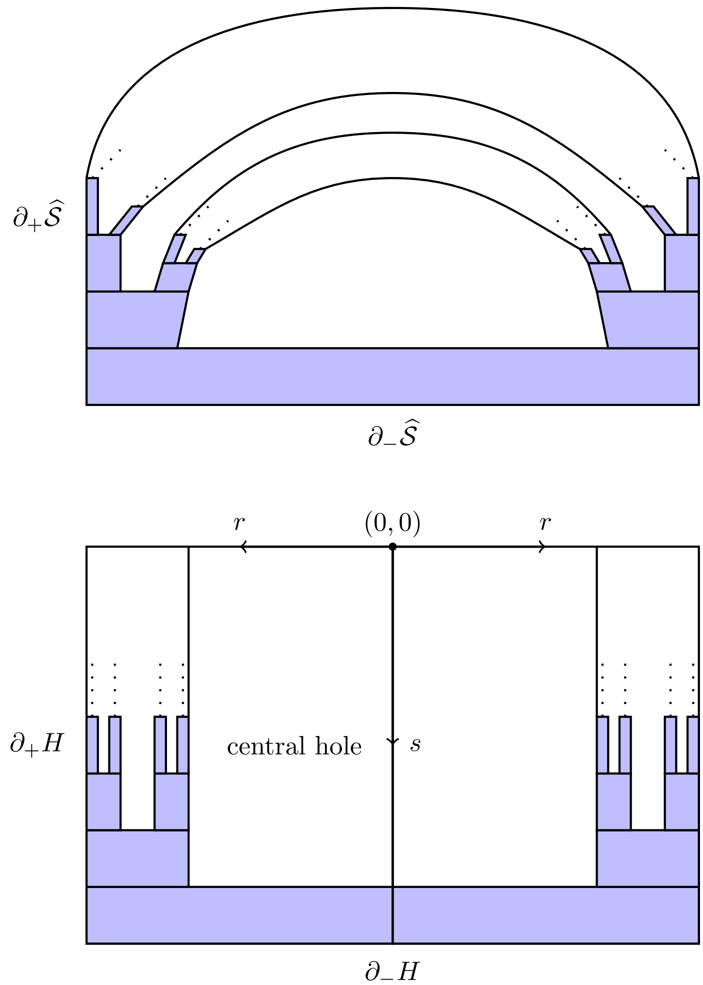


Figure 2.8. The design in a skyscraper $\widehat{\mathcal{S}}$ (top) and in the standard handle $H := D^2 \times D^2$ (bottom). The bottom picture uses coordinates $(s, \theta; r, \varphi)$, namely polar coordinates in the two factors of $D^2 \times D^2$. The θ coordinate is suppressed and only $\varphi = 0, \pi$ are shown.

Step 7 (Find a common subset of $\widehat{\mathcal{S}}$ and $D^2 \times D^2$, Sections 28.3 and 28.4). By the Skyscraper Embedding Theorem, we know that any two successive storeys of the given skyscraper $\widehat{\mathcal{S}}$ contains an embedded skyscraper with the correct attaching region. Split up the storeys of the skyscraper into consecutive pairs, and then apply the Skyscraper Embedding Theorem to each pair. Apply the same process – split into pairs of consecutive storeys and apply the Skyscraper Embedding Theorem – to each of the newly found skyscrapers, and iterate. In this manner we find uncountably many skyscrapers embedded within the original $\widehat{\mathcal{S}}$, all with the same attaching region. Since the vertical boundary of each embedded skyscraper is homeomorphic to $D^2 \times S^1$, we have also found uncountably many copies of $D^2 \times S^1$ inside $\widehat{\mathcal{S}}$. In Section 28.3, we use these vertical boundary solid tori as guides to fill up most of the space within $\widehat{\mathcal{S}}$ using collars within the finite storeys of the embedded skyscrapers. These collars get thinner as we climb up the storeys of any given skyscraper. The region of $\widehat{\mathcal{S}}$ filled up by these collars and the endpoints of the embedded skyscrapers is called *the design in the skyscraper* $\widehat{\mathcal{S}}$. The corresponding abstract space is called *the design*. Note that the design in the skyscraper contains the full $D^2 \times S^1$ vertical boundaries of the embedded skyscrapers, along with corresponding tapering collars.

Next, we need to embed the design in $D^2 \times D^2$. The image of this embedding is called *the design in the standard handle* $D^2 \times D^2$. We define this embedding in Section 28.4 using the Kirby diagrams for skyscrapers studied in Chapter 13. This embedding has the property that the solid tori corresponding to the vertical boundaries of embedded skyscrapers in $\widehat{\mathcal{S}}$ are mapped to $D^2 \times S^1_r \subseteq D^2 \times D^2$ for certain values of r, where $S^1_r \subseteq D^2$ is the circle of radius $r \in [0, 1]$. By definition there is a homeomorphism between the design in $\widehat{\mathcal{S}}$ and the design in $D^2 \times D^2$. The closures of the connected components of the complement of the design in $\widehat{\mathcal{S}}$ are called the *gaps* and those in $D^2 \times D^2$ are called the *holes*. The design in both a skyscraper and the standard handle are shown in Figure 2.8.

From the construction of the design, it will be straightforward to show that the quotient of $\widehat{\mathcal{S}}$ by the gaps is homeomorphic to the quotient of $D^2 \times D^2$ by the holes, but this by itself does not tell us very much. What we need to understand is the relationship between $\widehat{\mathcal{S}}$ and $\widehat{\mathcal{S}}/\text{gaps}$ and between $D^2 \times D^2$ and $D^2 \times D^2/\text{holes}$. This is exactly the purview of the field of decomposition space theory from Part I. In certain situations, results in decomposition space theory tell us that the quotient map collapsing a collection of subsets of a space X into distinct points (one for each subset) is *approximable by homeomorphisms*, which means that there is a sequence of homeomorphisms converging to the quotient map. In this situation, we say that the collection of subsets is *shrinkable* or that it *shrinks*. In particular, this means that the quotient space is homeomorphic to the original space. We have already seen an instance of this when we saw that the vertical boundary of a skyscraper is homeomorphic to $D^2 \times S^1$. If we could show that the collection of holes and the collection of gaps are shrinkable, we would have a sequence of homeomorphisms that would prove the theorem.

$$\text{Naïve wish:} \quad \widehat{\mathcal{S}} \cong \widehat{\mathcal{S}}/\text{gaps} \cong D^2 \times D^2/\text{holes} \cong D^2 \times D^2.$$

Unfortunately, it turns out that the collections of holes and of gaps are not shrinkable. Indeed, from our work with Kirby diagrams, we see that the holes in the standard $D^2 \times D^2$ are in fact copies of $S^1 \times D^3$, corresponding to neighbourhoods of the components of an iterated mixed ramified Bing-Whitehead link in a solid torus $(\times[0, 1])$. Thus the holes cannot shrink, since any shrinkable set must be simply connected. Luckily, we can make a reasonably elementary modification to the holes and gaps to make them shrinkable.

Step 8 (Modify the holes and gaps, Section 28.5). As noted, the holes in the standard $D^2 \times D^2$ are in fact copies of $S^1 \times D^3$, corresponding to neighbourhoods of the components of an iterated mixed ramified Bing-Whitehead link in a solid torus $(\times[0, 1])$. Recall that uncountably many solid tori of the form $D^2 \times S^1_r$ are contained in the design in the standard handle. Moreover, each hole has such a solid torus arbitrarily close to it. Inside these solid tori we can find embedded discs to cap off longitudes of each hole. These discs were christened *red blood cell discs* by Freedman, since suppressing the fourth dimension each hole with such a disc glued on looks like a red blood cell (see Figure 2.9).

Since the red blood cell discs are contained within the design in the standard handle, their interiors do not intersect the holes. The holes, after they have been

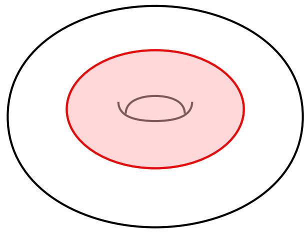


Figure 2.9. A red blood cell.

plugged with a disjoint collection of red blood cell discs, are called *holes*$^+$. Note that this entire process was possible since we had helpful solid tori close to the holes. This is another payoff of boundary shrinking, which does not hold for Casson handles.

Using the identification between the design in the skyscraper and the design in the standard handle, and the fact that the red blood cell discs are located inside the latter, we can find corresponding discs for the gaps within the design in the skyscraper. Once we cap off the gaps with these discs, we obtain the *gaps*$^+$.

Now the holes$^+$ and gaps$^+$ are at least simply connected, so we have some hope that they will shrink, and indeed they do.

$$\text{Informed desire:}\quad \widehat{\mathcal{S}} \cong \widehat{\mathcal{S}}/\text{gaps}^+ \cong D^2 \times D^2/\text{holes}^+ \cong D^2 \times D^2$$

relative to the attaching regions $S^1 \times D^2$. Showing that these three homeomorphisms exist is our remaining task.

Step 9 (Show that $D^2 \times D^2 \cong D^2 \times D^2/\text{holes}^+$, Proposition 28.21). We will show that the holes$^+$ shrink. This is called the α *shrink* and follows from the *Starlike Null Theorem*, which states that null, recursively starlike-equivalent decompositions shrink (Corollary 9.18). In order to apply this theorem, we show that the collection of holes$^+$ is *null*, meaning that only finitely many of the holes$^+$ have diameter larger than any given $\varepsilon > 0$, and that each hole$^+$ is *recursively starlike-equivalent*, meaning roughly that it is built out of starlike pieces.

Similarly to the case for holes and gaps, it is quite easy to see that $\widehat{\mathcal{S}}/\text{gaps}^+ \cong D^2 \times D^2/\text{holes}^+$. By the previous step, we now also know that $D^2 \times D^2/\text{holes}^+ \cong D^2 \times D^2$.

Step 10 (Show that $\widehat{\mathcal{S}}$ is homeomorphic to $\widehat{\mathcal{S}}/\text{gaps}^+$, Section 28.6.3). The last remaining step is to show that the gaps$^+$ shrink; this is called the β *shrink*. We use the *Ball to Ball Theorem* (Theorem 10.1), which gives a sufficient condition for a surjective map $f\colon D^4 \to D^4$ to be approximable by homeomorphisms. Recall that this means there are homeomorphisms arbitrarily close to the given surjective map, and moreover, these homeomorphisms agree with f on a collar of D^4. This might seem rather incongruous right now, since in order to apply this theorem to the quotient map by gaps$^+$, it seems like we would need to know that $\widehat{\mathcal{S}}$ is a ball, which is what we are trying to prove. We are able to apply the theorem without employing a circular argument by a clever trick. The *Collar Adding Lemma* (Lemma 25.1) tells us that if we add a collar to a skyscraper, then the resulting space is homeomorphic to a ball D^4. The more well known 3-dimensional analogue of this fact about the complement of the Alexander horned ball is discussed in Chapter 4. Extend the quotient map $\widehat{\mathcal{S}} \to \widehat{\mathcal{S}}/\text{gaps}^+$ via the identity on collars, using the fact that the total boundary of $\widehat{\mathcal{S}}$ is S^3, which follows from boundary shrinking.

We know now that both the domain and codomain of this extended map is homeomorphic to D^4. Then to apply the Ball to Ball Theorem, we need to show that the collection of gaps$^+$ is null and that their image in the quotient is nowhere dense. The former follows from our construction and the latter from straightforward point set topology. Then, the Ball to Ball Theorem applied to the extended map on collars yields a homeomorphism that agrees with the map on the added collars. In particular, throwing away the collars, such a homeomorphism restricts to a homeomorphism $\widehat{\mathcal{S}} \to \widehat{\mathcal{S}}/\text{gaps}^+ \cong D^2 \times D^2$, which completes the proof of the disc embedding theorem.

2.4. Reader's guide

We remind the reader one last time that Part I provides an introduction to decomposition space theory, proving the theorems that we need in the later proof. Part II describes how to build skyscrapers starting from the hypothesis of the disc embedding theorem. Part IV proves that every skyscraper is homeomorphic to the standard 2-handle, relative to the attaching region. Part III, an interlude describing applications of the disc embedding theorem, uses techniques from Part II but is not needed for the proof of the disc embedding theorem. Parts I and II are largely independent of each other. In Part IV, we primarily need three results from Part I: the starlike null theorem, shrinking of mixed ramified Bing-Whitehead decompositions, and the Ball to Ball Theorem. The reader who already knows decomposition space theory, or who is willing to accept these results, may skip Part I. The precise input that one needs from Parts I and II for the proof in Part IV is summarised in Chapter 26. It should also be possible, therefore, to read Part IV now, and refer backwards for results and definitions when needed.

Part I

Decomposition space theory

In the first part of this book, we give a streamlined account of the field of *decomposition space theory*, sometimes called *Bing topology*, optimised for its eventual use in the last part of the proof of the disc embedding theorem in Part IV. In Chapter 3 we present two proofs of the Schoenflies theorem in the topological category as an introduction to topological manifold theory. In Chapter 4 we lay out the foundations of decomposition space theory, stating and proving the central *Bing shrinking criterion*. In Chapter 5 we give a detailed investigation of the complement of the Alexander horned ball, which we call the *Alexander gored ball*. Using this we describe the *Bing decomposition* of S^3 and show that it shrinks. In Chapter 6 we present an example of a decomposition that does not shrink, arising as a variant of the Bing decomposition. The *Whitehead decomposition* from Chapter 7 is not directly used later, but was a key component of Freedman's original proof of the disc embedding theorem using Casson handles, and illustrates an interesting phenomenon that taking the product with $\mathbb{R}$ often converts a decomposition that does not shrink into one that does. In Chapter 8 we study *mixed ramified Bing-Whitehead decompositions* in detail. These will be central to our work in Part IV. In Chapter 9 we prove the *starlike null theorem*, which shows that null, recursively starlike-equivalent decompositions shrink and in Chapter 10 we prove the ball to ball theorem; the latter two are crucial for the final steps of the proof of the disc embedding theorem.

CHAPTER 3

The Schoenflies theorem after Mazur, Morse, and Brown

Stefan Behrens, Allison N. Miller, Matthias Nagel, and Peter Teichner

We introduce some basic techniques in the study of topological manifolds by means of a discussion of the Schoenflies theorem. First we present the proof of Mazur and Morse using the Eilenberg swindle and a technique called *push-pull*. These techniques exemplify the kind of arguments often used in the study of topological manifolds. Then we explain Brown's alternative proof of the Schoenflies theorem as an introduction to decomposition space theory, or *shrinking*.

The Schoenflies problem is a fundamental question about spheres embedded in Euclidean space. Denote the d-dimensional Euclidean space by $\mathbb{R}^d$, the closed unit disc or ball in $\mathbb{R}^d$ by D^d, and the d-dimensional sphere by S^d. We identify S^d with the boundary ∂D^{d+1}. The original Schoenflies problem can be stated as the conjecture that *for all d, every continuous embedding of S^d into $\mathbb{R}^{d+1}$ extends to a continuous embedding of D^{d+1} into $\mathbb{R}^{d+1}$.*

In 1913, the 1-dimensional case, more commonly known as the Jordan curve theorem, was proved in full generality by Caratheodory [**Car13**] and Osgood-Taylor [**OT13**] using elaborate methods from complex analysis. The 2-dimensional case was studied in the 1920s by Alexander who first circulated an unpublished manuscript claiming a proof but soon discovered a counterexample [**Ale24**], which is now called the *Alexander horned sphere*, shown in Figure 3.1. Later, Alexander found that the Schoenflies conjecture holds in dimension two given the existence of a *bicollar* [**Ale30**].

Definition 3.1. A continuous embedding $f\colon S^d \to \mathbb{R}^{d+1}$ has a *bicollar* if f extends to a continuous embedding $F\colon S^d \times [-1,1] \to \mathbb{R}^{d+1}$ such that F restricted to $S^d \times \{0\}$ is equal to f. We say that F is a *bicollared embedding* of S^d.

With the bicollared hypothesis added, the following became known as the Schoenflies conjecture.

Conjecture 3.2 (Schoenflies). *For all d, every bicollared embedding of S^d into $\mathbb{R}^{d+1}$ extends to a continuous embedding of D^{d+1} into $\mathbb{R}^{d+1}$.*

In the thirty years that followed almost no progress was made. In the 1950s there was pervasive pessimism among manifold topologists regarding the topological category. A watershed moment came in 1959 when Mazur gave his partial proof of the Schoenflies conjecture [**Maz59**], which we now explain.

3.1. Mazur's theorem

Mazur's proof uses a principle known as the *Eilenberg swindle*, which appears, for example, in the proof of the following observation in commutative algebra. Let A

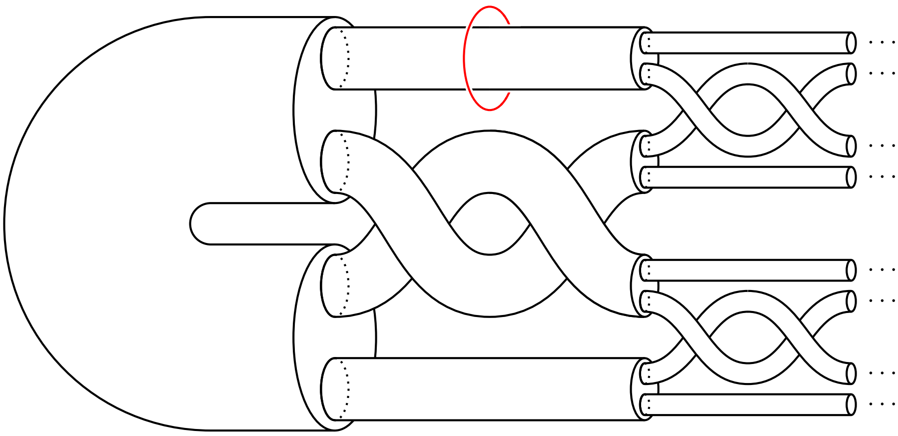


Figure 3.1. The Alexander horned sphere. Perform the indicated infinite construction, then add in a Cantor set to compactify the union of the tubes and obtain a topological embedding of a 2-sphere in $\mathbb{R}^3$. The complement of the interior region is not simply connected since, for example, the red circle is not null-homotopic.

be any projective module over some ring. Since A is projective, it can be written as a direct summand $A \oplus B \cong F$ where F is a free module and B is some module. Then on one hand we have

$$(A \oplus B) \oplus (A \oplus B) \oplus (A \oplus B) \oplus \cdots \cong F^\infty$$

while, on the other hand, a different grouping of the summands gives

$$A \oplus (B \oplus A) \oplus (B \oplus A) \oplus (B \oplus A) \oplus \cdots \cong A \oplus F^\infty$$

since the direct sum is associative and $B \oplus A \cong A \oplus B \cong F$. Thus, $F^\infty \cong A \oplus F^\infty$. In other words, A becomes an infinite dimensional free module upon direct sum with an infinite dimensional free module. That is, any projective module is *stably free* in the infinite dimensional context.

EXAMPLE 3.3 (Do knots have inverses?). The following is a standard application of the Eilenberg swindle in topology. Knots in $\mathbb{R}^3$ (or S^3) can be added by forming connected sums. We ask whether, given a knot A, there is a knot B such that the connected sum $A\#B$ is ambiently isotopic to the trivial knot U, that is, $A\#B \cong U$. We think of knots as strands within a cylinder, and indeed, every knot is the "braid

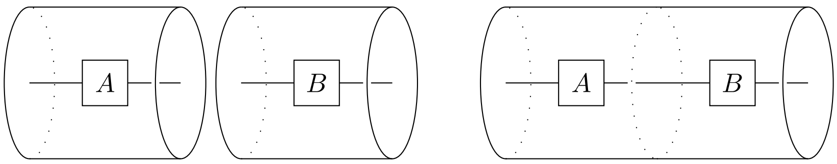


Figure 3.2. Adding knots in cylinders. The boxes denote tangles, that is, the braid closure of the strand lying within the box labelled A (respectively, B) is the knot A (respectively, B).

closure" of such a knotted strand. Then, the connected sum operation is realised by stacking cylinders next to each other, as shown in Figure 3.2. This operation is both commutative and associative. To see the commutativity, start with $A\#B$, shrink B so that it becomes very small compared to A, slide it along A to the other side, and let it grow again.

Assume that a knot A has an inverse B, that is $A\#B \cong U$. This implies we can unknot $A\#B$ using an ambient isotopy entirely supported in the two cylinders. Now the swindle works as follows. Take the connected sum of infinitely many copies of $A\#B$ and think of the resulting knot as living in a cone, which in turn lives in a cylinder, as shown in Figure 3.3. The cone forces the summands to get progressively smaller, so they limit to a point at the tip of the cone. Then we have an ambient

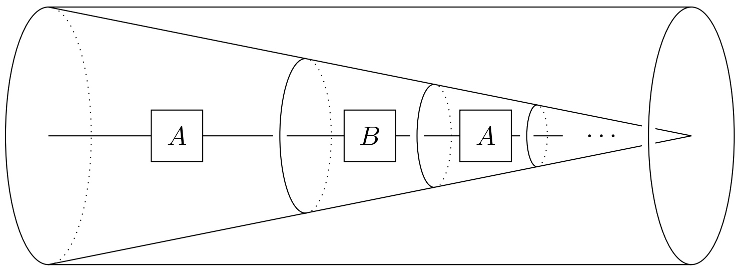


Figure 3.3. By stacking cylinders together, we construct the connected sum of infinitely many copies of $A\#B$ in a cone.

isotopy

$$(A\#B)\#(A\#B)\#(A\#B)\#\cdots \cong \#_{i=1}^{\infty}U \cong U$$

while a different grouping gives another ambient isotopy

$$A\#(B\#A)\#(B\#A)\#(B\#A)\#\cdots \cong A\#(\#_{i=1}^{\infty}U) \cong A$$

where we use the fact that $B\#A \cong U$ and apply infinitely many small ambient isotopies. Thus, A must be ambiently isotopic to the trivial knot. This proves that a nontrivial knot does not admit an inverse.

The above proof has the drawback that it loses category, that is we may have started with smooth or piecewise linear knots but the conclusion holds only in the topological category, since the ambient isotopy we constructed may not be smooth or piecewise linear at the cone point: we obtain a homeomorphism of S^3 sending A to the unknot, rather than a diffeomorphism. Other proofs of the non-cancellation of knots, such as the proof using additivity of the Seifert genus, do not have this drawback.

Mazur used the Eilenberg swindle to give a proof of the Schoenflies theorem, with a hypothesis about a *standard spot*.

Definition 3.4. Let $i\colon S^d \times [-1,1] \to \mathbb{R}^{d+1}$ be a bicollared embedding with a point $p \in S^d$ such that $i(p,0) = 0$. Write $\mathbb{R}^{d+1}$ as $\mathbb{R}^d \times \mathbb{R}$ and then the function i as (i_1, i_2) where $i_1\colon S^d \times [-1,1] \to \mathbb{R}^d$ and $i_2\colon S^d \times [-1,1] \to \mathbb{R}$.

We say that $(p,0) \in S^d \times [-1,1]$ is a *standard spot* of i if there is a standard d-dimensional disc $D^d \subseteq S^d$ around p such that

(a) the function i maps $D^d \times \{0\}$ to a standard round disc in $\mathbb{R}^d \times \{0\}$ and
(b) for each $q \in D^d$, the interval $\{q\} \times [-1,1]$ in $S^d \times [-1,1]$ is mapped by i such that $i(q,t) = (i_1(q,0), t)$.

Note that, in particular, the closure of the complement of D^d in S^d is also a standard disc. Morally, this definition means that i is "as standard as possible" around p.

Theorem 3.5 (Mazur [**Maz59**]). *Let $d \geq 1$ and let $i\colon S^d \times [-1,1] \to \mathbb{R}^{d+1}$ be a bicollared embedding with a standard spot. Then i extends to a continuous embedding of D^{d+1}.*

Proof. Let $i\colon S^d \times [-1,1] \to \mathbb{R}^{d+1}$ be the given bicollared embedding with a standard spot $(p,0)$. By passing to the one-point compactification of $\mathbb{R}^{d+1}$ we can consider i to be an embedding $S^d \times [-1,1] \hookrightarrow S^{d+1}$. Let $D^d \subseteq S^d$ be the disc in the definition of the standard spot. Cut out the image $i(D^d \times [-\frac{1}{2}, \frac{1}{2}]) = i(D^d) \times [-\frac{1}{2}, \frac{1}{2}]$ around $i(p,0)$. By definition, we have removed a standard ball from S^{d+1}, so the closure of the complement is also a standard ball (in particular, this does not assume the Schoenflies theorem).

Next, we claim that the space $\overline{S^{d+1} \smallsetminus i(S^d \times [-\frac{1}{2}, \frac{1}{2}])}$ has two components, as indicated in Figure 3.4 in the case $d = 1$. To see this, let $X := i(S^d \times [-\frac{1}{2}, \frac{1}{2}])$ and let $Y := \overline{S^{d+1} \smallsetminus X}$. Then $S^{d+1} = X \cup_{i(S^d \times \{-\frac{1}{2}, \frac{1}{2}\})} Y$, and so the Mayer-Vietoris sequence yields

$$H_1(S^{d+1}) \to H_0\big(S^d \times \big\{ -\tfrac{1}{2}, \tfrac{1}{2}\big\}\big) \to H_0(X) \oplus H_0(Y) \to H_0(S^{d+1}) \to 0.$$

To apply the Mayer-Vietoris sequence, we use that $i(S^d \times \{\pm\frac{1}{2}\})$ sits inside a larger collar, so is itself bicollared. Since $d \geq 1$, we have that $H_1(S^{d+1}) = 0$ and $\mathbb{Z} \cong H_0(S^{d+1}) \cong H_0(S^d) \cong H_0(X)$. We compute that $H_0(Y) \cong \mathbb{Z}^2$ and so Y has two connected components as claimed. We call these two pieces A^+ and A^-, where A^- is the piece contained in $\mathbb{R}^{d+1} \subseteq S^{d+1}$.

We also see from the existence of the standard spot that the boundary of $A^\pm$ is a d-dimensional sphere that is decomposed into two standard d-dimensional discs $P^\pm$ and $Q^\pm$, as shown in Figure 3.4, where $P^+ = i(D^d \times \{-\frac{1}{2}\})$, $P^- = i(D^d \times \{\frac{1}{2}\})$, and $Q^\pm$ are the closures of the complementary regions.

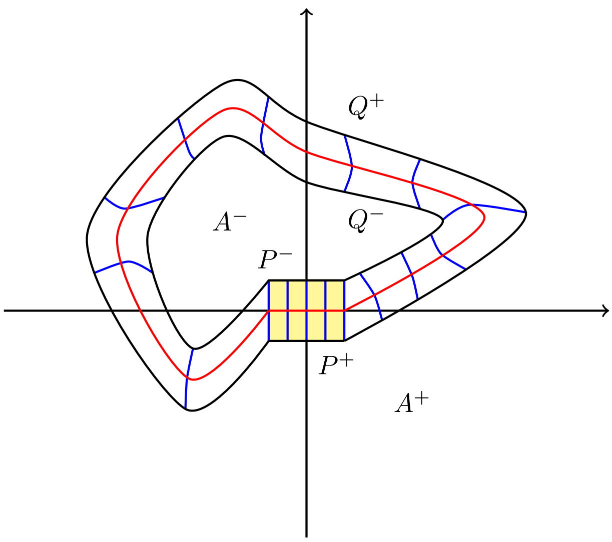


Figure 3.4. Mazur's partial proof of the Schoenflies conjecture. Red denotes the image $i(S^d)$. The standard spot is shown around the origin. Let $A^\pm$ denote the connected components of the complement of $i(S^d \times [-\frac{1}{2}, \frac{1}{2}])$. The boundary $i(S^d \times \{-\frac{1}{2}\})$ is decomposed into $P^+ \cup Q^+$ and $i(S^d \times \{\frac{1}{2}\})$ is decomposed into $P^- \cup Q^-$.

Consider the space

$$A^- \cup_Q A^+ := A^- \cup_{Q^-} i((\overline{S^d \smallsetminus D^d}) \times [-\tfrac{1}{2}, \tfrac{1}{2}]) \cup_{Q^+} A^+.$$

By definition, this is the closure of the complement in S^{d+1} of $i(D^d \times [-\frac{1}{2}, \frac{1}{2}])$, and we have already established that it is homeomorphic to D^{d+1}. Next we show that the space

$$A^- \cup_P A^+ := A^- \cup_{P^-} i(D^d \times [-\tfrac{1}{2}, \tfrac{1}{2}]) \cup_{P^+} A^+$$

is also homeomorphic to D^{d+1}. To see this, note that $P^\pm$ and $Q^\pm$ are ambiently isotopic in $\partial A^\pm$, via some ambient isotopy

$$F^\pm \colon \partial A^\pm \times [0,1] \to \partial A^\pm,$$

with $F^\pm|_{\partial A^\pm \times \{0\}} = \mathrm{Id}$, since $P^\pm$ and $Q^\pm$ may be considered to be the standard northern and southern hemispheres of the d-dimensional sphere $\partial A^\pm$. By construction, $\partial A^\pm$ is collared. Thus there is an embedding $\partial A^\pm \times [0,1] \to A^\pm$ with $\partial A^\pm \times \{1\}$ mapped homeomorphically to $\partial A^\pm$. Then, we have the following homeomorphism obtained by inserting the ambient isotopies into the boundary collars.

$$\begin{tikzcd}
A^- \arrow[d, "\mathrm{Id}"] & \cup_{P^-} & \partial A^- \times [0,1] \arrow[d, "F^- \times \mathrm{Id}"] & \cup_{P^-} & i(D^d \times [-\frac{1}{2}, \frac{1}{2}]) \arrow[d, "g"] & \cup_{P^+} & \partial A^+ \times [0,1] \arrow[d, "F^+ \times \mathrm{Id}"] & \cup_{P^+} & A^+ \arrow[d, "\mathrm{Id}"] \\
A^- & \cup_{P^-} & \partial A^- \times [0,1] & \cup_{Q^-} & i((\overline{S^d \smallsetminus D^d}) \times [-\frac{1}{2}, \frac{1}{2}]) & \cup_{Q^+} & \partial A^+ \times [0,1] & \cup_{P^+} & A^+
\end{tikzcd}$$

The middle map is obtained using the abstract homeomorphism $D^d \cong \overline{S^d \smallsetminus D^d}$. The diagram shows that $A^- \cup_P A^+ \cong A^- \cup_Q A^+ \cong D^{d+1}$ as desired.

We are now ready for the Eilenberg swindle. We have the following sequence of homeomorphisms.

$$\begin{aligned}
D^{d+1} &\cong \left(A^- \cup_Q A^+\right) \cup_{P^-} i(D^d \times [-\frac{1}{2}, \frac{1}{2}]) \cup_{P^+} \left(A^- \cup_Q A^+\right) \cup_{P^-} \cdots \cup \{\infty\} \\
&\cong A^- \cup_{Q^-} i((\overline{S^d \smallsetminus D^d}) \times [-\frac{1}{2}, \frac{1}{2}]) \cup_{Q^+} \left(A^+ \cup_P A^-\right) \cup_{Q^-} \cdots \cup \{\infty\} \\
&\cong A^- \cup_{Q^-} i((\overline{S^d \smallsetminus D^d}) \times [-\frac{1}{2}, \frac{1}{2}]) \cup_{Q^+} D^{d+1} \cup_{Q^-} D^{d+1} \cdots \cup \{\infty\} \\
&\cong A^- \cup_{Q^-} i((\overline{S^d \smallsetminus D^d}) \times [-\frac{1}{2}, \frac{1}{2}]) \cup_{Q^+} D^{d+1} \\
&\cong A^-.
\end{aligned}$$

For the first homeomorphism above, we are using the fact that $A^- \cup_Q A^+ \cong D^{d+1}$ and that gluing infinitely many balls in pairs along balls of one lower dimension in their boundaries and then taking the one-point compactification gives another ball. The second step is the Eilenberg swindle, where the re-bracketing occurs. The third step uses that $A^+ \cup_P A^- \cong D^{d+1}$, as shown above. Then we use again the fact that a compactified infinite sequence of balls glued together along balls of one lower dimension is homeomorphic to a ball. The last homeomorphism is easier: the boundary connected sum of finitely many balls is homeomorphic to a ball, and since ∂A^- is collared, boundary connected sum with a ball along part of its boundary is trivial. We have now shown that the space A^- is homeomorphic to D^{d+1}. Note that the closure of the component of the complement of $i(S^d \times \{0\})$ in $\mathbb{R}^{d+1} \subseteq S^{d+1}$ is A^- equipped with a boundary collar and thus is also homeomorphic to D^{d+1}. The proof is then completed by the Alexander trick, which provides an extension of a given homeomorphism $S^d \to S^d$ to a homeomorphism $D^{d+1} \to D^{d+1}$. $\square$

Note that we could reverse the rôles of A^- and A^+ in the proof above to conclude that A^+ is also a ball. Thus, we have shown that given a bicollared embedding $i\colon S^d \hookrightarrow S^{d+1}$ with a standard spot, both of the connected components of the complement $S^{d+1} \smallsetminus i(S^d)$ have closures homeomorphic to D^{d+1}. Next we show that the standard spot is not required.

3.2. Morse's theorem

Mazur's work generated a lot of interest in the problem of removing the standard spot hypothesis. This was solved in 1960 in a paper by Morse [**Mor60**] using a technique called *push-pull*. We introduce it by proving a theorem that uses the technique.

THEOREM 3.6 (Application of push-pull). *Let X and Y be compact metric spaces. If $X \times \mathbb{R}$ is homeomorphic to $Y \times \mathbb{R}$, then $X \times S^1$ is homeomorphic to $Y \times S^1$.*

PROOF. Let $h\colon X \times \mathbb{R} \to Y \times \mathbb{R}$ be a homeomorphism. The key point in this argument will be that $Y \times \mathbb{R}$ has two product structures, the intrinsic one and the one induced from $X \times \mathbb{R}$ via h.

Let X_t denote $X \times \{t\}$ for $t \in \mathbb{R}$ and let $X_{[t,u]}$ denote $X \times [t,u]$ for $[t,u] \subseteq \mathbb{R}$. Similarly, let Y_s denote $Y \times \{s\}$ for $s \in \mathbb{R}$ and let $Y_{[r,s]}$ denote $Y \times [r,s]$ for $[r,s] \subseteq \mathbb{R}$. By compactness of X and Y, there exist $a < c < e$ and $b < d$ such that

(1) Y_a, Y_c, Y_e, $h(X_b)$, and $h(X_d)$ are pairwise disjoint in $Y \times \mathbb{R}$,
(2) $h(X_b) \subseteq Y_{[a,c]}$,
(3) $Y_c \subseteq h(X_{[b,d]})$, and
(4) $h(X_d) \subseteq Y_{[c,e]}$,

as illustrated in the leftmost panel in Figure 3.5. This may be achieved by first fixing a, and then choosing as follows.

- Choose b so that (1) is satisfied for a and b.
- Choose $c > a$ so that (1) and (2) are satisfied for a, b, and c.
- Choose $d > b$ so that (1) and (3) are satisfied for a, b, c, and d.
- Choose $e > c$ so that (1) and (4) are satisfied.

Now we construct a self-homeomorphism χ of $Y \times \mathbb{R}$ as the composition

$$\chi = C^{-1} \circ P_Y \circ P_X \circ C,$$

where the steps are illustrated in Figure 3.5. The maps P_X and P_Y will constitute the actual pushing and pulling while C, which we might call *cold storage*, makes sure that nothing is pushed or pulled unless it is supposed to be.

The maps are obtained as follows:

- The map C rescales the intrinsic $\mathbb{R}$-coordinate of $Y \times \mathbb{R}$ such that $C(Y_{[a,c]})$ lies below $h(X_b)$ and leaves $h(X_d)$ untouched. We require C to be the identity on $Y_{[c+\varepsilon,\infty)}$ and $Y_{(-\infty,a]}$, for ε small enough so that $Y_{c+\varepsilon} \subsetneq h(X_{[b,d]})$.
- The map P_X pushes $h(X_d)$ down to $h(X_b)$ along the $\mathbb{R}$-coordinate induced by h, that is, the image of the product structure of $X \times \mathbb{R}$, without moving $C(Y_{[a,c]})$.
- The map P_Y pulls $h(X_b) = (P_X \circ C \circ h)(X_d)$ up along the intrinsic $\mathbb{R}$-coordinate of $Y \times \mathbb{R}$ so that it lies above the support of C^{-1}, again without moving $C(Y_{[a,c]})$. This can be done in such a way that P_Y is supported below Y_e.

The map χ is the identity outside of $Y_{[a,e]}$. Observe that χ leaves $h(X_b)$ untouched and that $\chi(h(X_d))$ appears as a translate of $h(X_b)$ in the intrinsic $\mathbb{R}$-coordinate.

By repeating the above process infinitely many times we construct a homeomorphism $\overline{\chi}\colon Y \times \mathbb{R} \to Y \times \mathbb{R}$ so that $H := \overline{\chi} \circ h\colon X \times \mathbb{R} \to Y \times \mathbb{R}$ is periodic. Here, in the

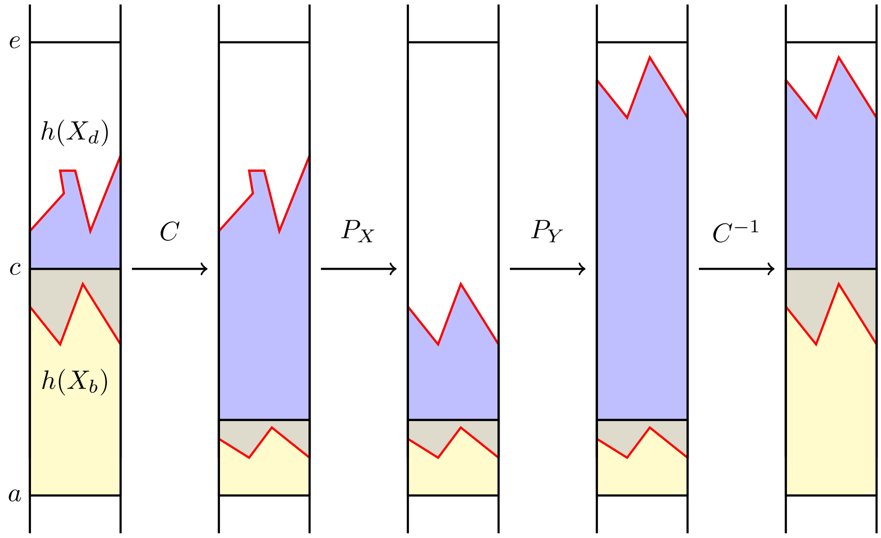


Figure 3.5. The push-pull construction. Each panel depicts the space $Y \times \mathbb{R}$. The blue and yellow regions denote $h(X_{[b,d]})$ and $Y_{[a,c]}$, respectively. Note that the regions overlap.

construction of $\overline{\chi}$ we use the fact that χ has support in $[a, e]$ and thus, each point in $Y \times \mathbb{R}$ is moved finitely many times in the definition of $\overline{\chi}$. Therefore we may compose infinitely many homeomorphisms in this case. The periodic homeomorphism H induces a homeomorphism $X \times S^1 \to Y \times S^1$ as desired. $\square$

Remark 3.7. The converse of Theorem 3.6 is not true in general. There exist compact manifolds X and Y such that $X \times S^1$ and $Y \times S^1$ are diffeomorphic but $X \times \mathbb{R}$ and $Y \times \mathbb{R}$ are not even homotopy equivalent [**Cha65**, Theorem 3.9].

Remark 3.8. The compactness hypothesis of Theorem 3.6 is necessary. That is, there exist examples of noncompact metric spaces X and Y such that $X \times \mathbb{R}$ and $Y \times \mathbb{R}$ are homeomorphic but $X \times S^1$ and $Y \times S^1$ are not, as follows. Let $\Sigma_{g,n}$ denote the compact, orientable surface with genus g and n boundary components. Note that $\Sigma_{g,1} \times [0, 1]$ is homeomorphic to $\Sigma_{0,2g+1} \times [0, 1]$. Indeed, both are obtained from D^3 by attaching $2g$ orientable 1-handles, and there is an essentially unique way to attach orientable 3-dimensional 1-handles to D^3. Let X and Y be the interiors of $\Sigma_{g,1}$ and $\Sigma_{0,2g+1}$, respectively. Then $X \times \mathbb{R}$ and $Y \times \mathbb{R}$ are homeomorphic (indeed, diffeomorphic) since they are the interiors of the homeomorphic spaces $\Sigma_{g,1} \times [0, 1]$ and $\Sigma_{0,2g+1} \times [0, 1]$, respectively. However, the end of $X \times S^1$ is homotopy equivalent to a torus, but the set of ends of $Y \times S^1$ is homotopy equivalent to the disjoint union of $2g + 1$ copies of tori. Therefore $X \times S^1$ is not homeomorphic to $Y \times S^1$.

Morse used the technique of push-pull to prove the following theorem.

Theorem 3.9 (Morse [**Mor60**]). *For all d, every bicollared embedding $S^d \times [0, 1] \to \mathbb{R}^{d+1}$ has a standard spot after applying a self-homeomorphism of $\mathbb{R}^{d+1}$.*

Proof (sketch). Consider a bicollared embedding $i \colon S^d \times [-1, 1] \to \mathbb{R}^{d+1}$ and fix a point $p \in S^d$. Up to translation, we can assume that $i(p, 0) = 0$. Choose local coordinates on a standard disc $D^d \subseteq S^d$ containing p, which yields an induced local coordinate system on $i(D^d \times [-1, 1]) \subseteq \mathbb{R}^{d+1}$, as shown in Figure 3.6.

In this new local coordinate system on $\mathbb{R}^{d+1}$, the embedded sphere has a standard spot, so it remains to extend it to a global coordinate system. We achieve this by

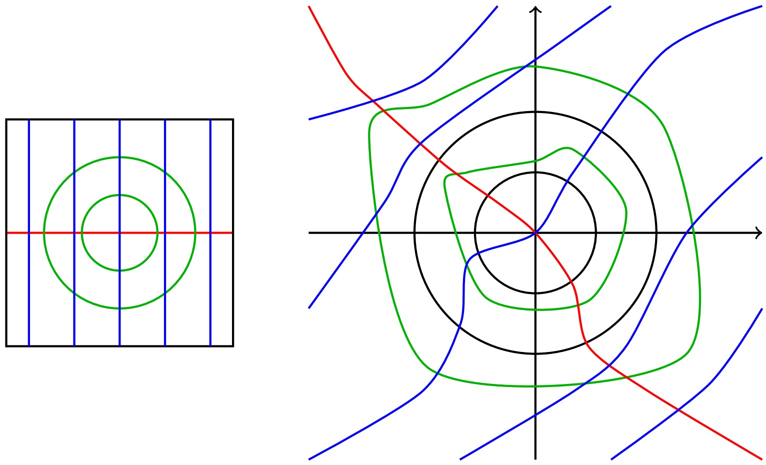


Figure 3.6. Creating a standard spot. On the left, the disc D^d appears as the central horizontal red segment. The blue vertical lines show the induced coordinate system on $D^d \times [0,1]$. A collection of standard round spheres are indicated in green. On the right, we show the image of $D^d \times [0,1]$ in $\mathbb{R}^{d+1}$. The proof of Theorem 3.9 compares the intrinsic round spheres (black) with the induced round spheres (green).

using a push-pull argument. The idea is to compare the standard polar coordinate system in $\mathbb{R}^{d+1}$ with the one induced by i. Again, by compactness we can find interlaced pairs of standard d-dimensional spheres in $\mathbb{R}^{d+1}$ and homeomorphically mapped spheres, as indicated in Figure 3.6. Then by using push-pull, we can find an isotopy that transforms one of the homeomorphically mapped spheres into a translate of the other homeomorphically mapped sphere along the standard radial coordinate and preserves a neighbourhood of the origin. Then extend the local chart by periodicity to cover all of $\mathbb{R}^{d+1}$. $\square$

Combining the results of Mazur and Morse we immediately deduce the following theorem.

Theorem 3.10 (Schoenflies theorem). *Let $d \geq 1$. Every bicollared embedding of S^d into $\mathbb{R}^{d+1}$ extends to a continuous embedding of D^{d+1}.*

As a historical note, by the time Morse had augmented Mazur's argument with his theorem, Brown had already given an independent and complete proof of the Schoenflies theorem, which we will discuss shortly.

We observe that the utility of the push-pull technique is in gaining control over a homeomorphism in one linear direction. As we will see, a major technical problem when working with topological manifolds is to gain control of a homeomorphism in many directions simultaneously. Results in this direction culminated in Kirby's work on the torus trick [**Kir69**].

We end this section by stating some more applications of push-pull to topological manifolds.

Theorem 3.11 ([**Bro62**, Theorem 3]). *A locally bicollared codimension one embedding in any topological manifold is globally bicollared.*

Theorem 3.12 ([**Bro62**, Theorem 2]). *The boundary of every topological manifold is collared.*

Theorem 3.13 ([**Arm70**, Theorem 2]). *Any two locally flat collars for either a codimension one submanifold or for the boundary of a topological manifold are ambiently isotopic to one another.*

3.3. Shrinking cellular sets

At the end of this chapter, we will give Brown's alternative proof of the Schoenflies theorem. In this section, we set the stage by introducing certain elementary notions from *decomposition space theory*, a field of ideas that will be central to the proof of the disc embedding theorem. In this section we follow [**Bro60**] and [**Dav07**].

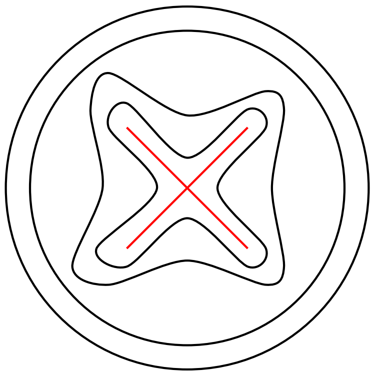


Figure 3.7. A cellular set (red) in D^2. The boundaries of embedded discs (black) converging to the cellular set are shown.

Definition 3.14. Let M^d be a d-dimensional manifold. A subset $X \subseteq M^d$ is said to be *cellular* if it is the intersection of countably many nested closed balls in M^d, that is if there exist embedded, closed d-dimensional balls $B_i \subseteq M^d$, $i \geq 1$, with $B_i \cong D^d$, such that $B_{i+1} \subseteq \operatorname{Int} B_i$ and $X = \bigcap_{i=1}^{\infty} B_i$.

Figure 3.7 illustrates that the letter

$$\mathsf{X} := \{(x,y) \in D^2 \mid x^2 = y^2, |x| \leq 1/2\}$$

is a cellular subset of D^2. Most of the cellular sets in this section will be denoted by the symbol X. We begin with some elementary properties of cellular sets.

Proposition 3.15. *Every cellular subset X of a manifold M is closed and compact.*

Proof. Let $\{B_i\}$ be the nested balls as in Definition 3.14. Then $X = \bigcap_{i=1}^{\infty} B_i$ is closed as an intersection of closed sets. Further X is compact since it is a closed subset of the compact space B_1. $\square$

Proposition 3.16. *Let X be a cellular set in a d-dimensional manifold M and let U be an open set with $X \subseteq U$. Then there exist embedded, closed d-dimensional balls $B_i \subseteq U$, $i \geq 1$ with $B_i \cong D^d$ and $B_{i+1} \subseteq \operatorname{Int} B_i$ for all i such that $X = \bigcap_{i=1}^{\infty} B_i$.*

Proof. By definition there exist embedded, closed d-dimensional balls $B_i \subseteq M$, $i \geq 1$, with $B_i \cong D^d$, such that $B_{i+1} \subseteq \operatorname{Int} B_i$ and $X = \bigcap_{i=1}^{\infty} B_i$. It suffices to show that there exists a j such that $B_j \subseteq U$. Suppose not. Then for all i, $B_i \cap (M \smallsetminus U)$ is nonempty. Choose a point $x_i \in B_i \cap (M \smallsetminus U) \subseteq B_1$ for each i. Since B_1 is sequentially compact, the sequence $\{x_i\}$ has a convergent subsequence $\{x_{i_k}\}$, converging to some $x \in B_1$. We assert that $x \in \bigcap B_i$. To see this, note that

$x_{i_k} \in B_{i_k}$ for all k. Fix ℓ. Then for $k \geq \ell$, $x_{i_k} \in B_{i_k} \subseteq B_{i_\ell}$. Since B_{i_ℓ} is closed, $x \in B_{i_\ell}$. Thus $x \in B_{i_\ell}$ for all ℓ, so $x \in \bigcap B_i$ as asserted. We therefore have that $x \in \bigcap B_i = X \subseteq U$ and $\{x_{i_k}\}$ is contained in the closed set $M \setminus U$, which is a contradiction since closed sets contain their limit points. □

REMARK 3.17. Note that cellularity is not an intrinsic property of a space X but rather depends on its specific embedding within the ambient space. For example, there exist non-cellular embeddings of a closed arc in S^3 [**Edw80**].

The key property of cellular sets is that they can be *shrunk* by homeomorphisms, as seen in the following proposition.

PROPOSITION 3.18. *Let X be a cellular set in a d-dimensional manifold M and let U be an open set with $X \subseteq U$. For every $\varepsilon > 0$ there exists a homeomorphism $h_\varepsilon \colon M \to M$ such that h_ε is the identity outside U and $\operatorname{diam} h_\varepsilon(X) < \varepsilon$.*

PROOF. By Proposition 3.16, there is a a closed ball B in U such that $X \subseteq \operatorname{Int} B$. Since X is closed by Proposition 3.15, there exists a collar N of ∂B disjoint from X. Find a ball D in $B \setminus N$ such that $\operatorname{diam} D < \varepsilon$. Now pick a homeomorphism $s_\varepsilon \colon B \to B$ which is the identity on the boundary ∂B and maps the complement of the collar N into D. Consequently, $s_\varepsilon(X) \subseteq D$ and therefore $\operatorname{diam} s_\varepsilon(X) < \varepsilon$. The map h_ε is obtained by extending s_ε to all of M by the identity map. □

For a homeomorphism h_ε as in the statement above, we say that $h_\varepsilon \colon M \to M$ *shrinks X in U to diameter less than ε.*

Our eventual goal is to use decomposition space theory, specifically the idea of shrinking, to approximate certain functions by homeomorphisms. Next we define precisely what this means.

Let X and Y be compact metric spaces. Recall that the uniform metric is defined by setting $d(f,g) = \sup_{x \in X} d_Y(f(x), g(x))$ for functions $f, g \colon X \to Y$. We denote the metric space of continuous functions from X to Y, equipped with the uniform metric, by $\mathcal{C}(X,Y)$. This is known to be a complete metric space [**Mun00**, Theorems 43.6 and 45.1]. Observe that the metric space $\mathcal{C}(X,X)$ contains the subspace $\mathcal{C}_A(X,X)$ of functions $f \colon X \to X$ with $f|_A = \operatorname{Id}_A$, for any subset $A \subseteq X$. This is a closed set in $\mathcal{C}(X,X)$ and thus is itself a complete metric space under the induced metric. The following definition formalises the notion of approximating functions by homeomorphisms.

DEFINITION 3.19. Let X and Y be compact metric spaces and let $f \colon X \to Y$ be a surjective continuous map. The map f is said to be *approximable by homeomorphisms* if there is a sequence of homeomorphisms $\{h_n \colon X \to Y\}_{n=1}^{\infty}$ that converges to f in $\mathcal{C}(X,Y)$.

In particular a necessary condition for f to be approximable by homeomorphisms is that X and Y are homeomorphic. In applications, frequently we will not know that X and Y are homeomorphic until we have shown that a map $f \colon X \to Y$ is approximable by homeomorphisms.

We will often wish to approximate quotient maps by homeomorphisms. This will only be meaningful when the quotient spaces are metric spaces. We record the following fact for use in this chapter. This will later be subsumed by Corollary 4.13.

Given a surjective map $f \colon X \to Y$ between topological spaces, we say that a subset $C \subseteq X$ is *saturated* (with respect to f) if whenever $f^{-1}(y)$ intersects C, for some $y \in Y$, we have $f^{-1}(y) \subseteq C$, or in other words the set C is a union of fibres of f.

Proposition 3.20. *Let M be a compact d-dimensional manifold, possibly with nonempty boundary. Let $X \subseteq \operatorname{Int} M$ be a cellular set. Then the quotient M/X is a compact metric space.*

Proof. Fix some metric on M inducing its topology. The quotient M/X is compact since M is compact. We show that M/X is Hausdorff. Let x denote the image of X in M/X. Choose $y, z \in M/X$ with $y \neq z$. Consider the quotient map $\pi\colon M \to M/X$. The restriction of π to the saturated open set $M \setminus X$ is an open, continuous bijection and thus a homeomorphism. If $y, z \neq x$, then $\pi^{-1}(y)$ and $\pi^{-1}(z)$ are distinct points in $M \setminus X$ with disjoint open neighbourhoods in $M \setminus X$ which are mapped by π to disjoint open neighbourhoods in M/X. Moreover, since M is a metric space, we can find disjoint open neighbourhoods of X and $\pi^{-1}(y)$ in M. These are saturated open sets and are thus mapped to (disjoint) open sets in M/X separating x and y. Therefore M/X is Hausdorff. This finishes the proof since the continuous image of a compact metric space in a Hausdorff space is metrisable [**Wil70**, Corollary 23.2, p. 166]. □

A simple class of quotient maps consists of those where a unique point in the codomain has more than one point in its pre-image. In other words, such a map is many to one on this pre-image but one to one everywhere else. The following terminology will be helpful in describing such maps.

Definition 3.21. Let $f\colon X \to Y$ be a map between topological spaces. The set $f^{-1}(y)$, where $y \in Y$, is called an *inverse set* of f if $|f^{-1}(y)| > 1$.

The following proposition shows that crushing a cellular set to a point does not change the homeomorphism type of a manifold.

Proposition 3.22. *Let M be a compact d-dimensional manifold, possibly with nonempty boundary. Let $X \subseteq \operatorname{Int} M$ be a cellular set. Then the quotient map $\pi\colon M \to M/X$ is approximable by homeomorphisms. In particular, the quotient space M/X is homeomorphic to M.*

Before giving the proof, we recall the following elementary lemma, since we will use it frequently.

Lemma 3.23 (Closed map lemma). *Every continuous map from a compact space A to a Hausdorff space B sends closed subsets of A to closed subsets of B.*

Proof. Let $U \subseteq A$ be closed. Then U is compact as a closed set in a compact space. Continuous maps preserve compactness, so $f(U)$ is compact. Finally, a compact subset of a Hausdorff space is closed, so $f(U) \subseteq B$ is closed. □

Proof of Proposition 3.22. Fix metrics on M and M/X, using Proposition 3.20. Most of the proof will consist of building a surjective, continuous function $f\colon M \to M$ which has X as its unique inverse set.

Since X is cellular, there is a family $\{B_i\}$ of closed balls with $B_0 \subseteq \operatorname{Int} M$, $B_{i+1} \subseteq \operatorname{Int} B_i$ for all i, and $\bigcap_i B_i = X$. We inductively define a family of homeomorphisms $f_i\colon M \to M$, starting with $f_0 = \mathrm{Id}_M$. Assume that f_i is already defined for some i. From the proof of Proposition 3.18, there is a homeomorphism $h_i\colon M \to M$ shrinking $f_i(B_{i+1})$ in $f_i(B_i)$ to diameter less than $\frac{1}{i+1}$, that restricts to the identity outside $\operatorname{Int} f_i(B_i)$. Define $f_{i+1} = h_i \circ f_i$. Note that $\operatorname{diam} f_i(B_i) < \frac{1}{i}$ for all i by construction.

Next we will show that the sequence $\{f_i\}$ in the complete metric space $\mathcal{C}_{\partial M}(M, M)$ is Cauchy. Fix integers $m > n$. Note that $f_m = f_n$ outside B_n. For every point $x \in B_n$, we have that $f_m(x), f_n(x) \in f_n(B_n)$ and as $\operatorname{diam} f_n(B_n) < \frac{1}{n}$, we get $d(f_m(x), f_n(x)) < \frac{1}{n}$. This implies that $d(f_n, f_m) < \frac{1}{n}$ in $\mathcal{C}_{\partial M}(M, M)$ and so

$\{f_i\}$ is a Cauchy sequence. We define f to be the limit of the sequence $\{f_i\}$, which exists since $\mathcal{C}_{\partial M}(M,M)$ is complete. By construction if $x \notin B_i$, then $f(x) = f_i(x)$.

Next, we show that f has the correct inverse sets. Let $z \in M$ be such that $z \notin B_i$ for some i, and let $x \in X$. Then

$$d(f(z), f(x)) = d(f_i(z), f(x)) \geq d(f_i(z), f_i(B_{i+1})) > 0.$$

Above, in the penultimate inequality, we use the fact that $f(x) \in f_i(B_{i+1})$. In the final inequality, we use that $B_{i+1} \subseteq \operatorname{Int} B_i$, so for every $z \notin B_i$, $d_X(z, B_{i+1}) > 0$. The inequality follows since f_i is a homeomorphism on $M \smallsetminus \operatorname{Int} B_{i+1}$. Thus $f(X)$ is disjoint from $f(M \smallsetminus X)$. Additionally, note that $\operatorname{diam} f_i(X) < \frac{1}{i}$ for all i and thus $f(X)$ consists of a single point y. As a result, $f^{-1}(y) = X$.

Next we show that $f^{-1}(z)$ for $z \neq y$ consists of precisely one element. Note that $f^{-1}(z) \subseteq M \smallsetminus X$. Thus $f^{-1}(z) = (f|_{M \smallsetminus X})^{-1}(z)$ and it suffices to show that $f|_{M \smallsetminus X}$ is injective. Given any two points $p, q \in M \smallsetminus X$, there exists some i such that $p, q \notin B_i$. Then, $f(p) = f_i(p)$ and $f(q) = f_i(q)$. Since each f_i is a homeomorphism and therefore injective, this completes the proof that X is the unique inverse set of f.

Finally we are ready to investigate the quotient map $\pi\colon M \to M/X$ directly. Note that the surjective map f descends to a map $M/X \to M$ via the quotient map and we obtain a bijective continuous function $\overline{f}\colon M/X \to M$.

$$\begin{tikzcd} M \arrow[r, "f"] \arrow[d, "\pi"'] & M \\ M/X \arrow[ur, "\overline{f}"'] & \end{tikzcd}$$

By the closed map lemma (Lemma 3.23), $\overline{f}$ is a homeomorphism. Note that $f = \overline{f} \circ \pi$.

Given the sequence of homeomorphisms $\{f_i\colon M \to M\}$ converging to f, consider the functions

$$\{\overline{f}^{-1} \circ f_i\colon M \to M/X\}.$$

These are homeomorphisms since $\overline{f}^{-1}$ and f_i are. The sequence $\{\pi_i := \overline{f}^{-1} \circ f_i\}$ converges to

$$\overline{f}^{-1} \circ f = \overline{f}^{-1} \circ \overline{f} \circ \pi = \pi$$

as desired, since $\overline{f}^{-1}$ is uniformly continuous. □

There is no need to restrict ourselves to the case of a single cellular set. We will show next that any finite collection of cellular sets in a manifold can be crushed to individual points (one per cellular set) while preserving the homeomorphism type of the manifold. We will need the following proposition.

PROPOSITION 3.24. *Let $f\colon X \to Y$ and $g\colon Y \to Z$ be maps between compact metric spaces that are approximable by homeomorphisms. Then $g \circ f\colon X \to Z$ is also approximable by homeomorphisms.*

We will use the next elementary result from analysis [**Rud76**, Theorem 4.19] in the proof of the proposition, but also many times in the future, so we record it here.

THEOREM 3.25 (Heine-Cantor theorem). *Let Y be a metric space, let X be a compact metric space, and let $f\colon X \to Y$ be a continuous function. Then f is uniformly continuous.*

PROOF OF PROPOSITION 3.24. Let $\varepsilon > 0$. The Heine-Cantor theorem (Theorem 3.25) implies that since Y is compact and g is continuous, the function $g\colon Y \to$

Z is uniformly continuous. Thus there is a $\delta > 0$ such that $d_Z(g(y), g(y')) < \frac{\varepsilon}{2}$ whenever $y, y' \in Y$ are such that $d_Y(y, y') < \delta$.

Recall that for two functions $f, f' \colon X \to Y$, the uniform metric is defined by $d(f, f') := \sup_{x\in X} d_Y(f(x), f'(x))$. Similarly for functions $g, g' \colon Y \to Z$ we have $d(g, g') := \sup_{y\in Y} d_Z(f(y), f'(y))$. Let $\{f_n \colon X \to Y\}$ be a sequence of homeomorphisms converging to f and let $\{g_m \colon Y \to Z\}$ be a sequence of homeomorphisms converging to g. That is, there exists $N > 0$ such that $d(f, f_n) < \delta$ whenever $n \geq N$, and similarly there exists $M > 0$ such that $d(g, g_m) < \frac{\varepsilon}{2}$ whenever $m \geq M$.

Let L be the maximum of M and N. Then for every $x \in X$ and every $n \geq L$, $d_Y(f(x), f_n(x)) < \delta$. By the uniform continuity property,

$$d_Z(g(f(x)), g(f_n(x))) < \tfrac{\varepsilon}{2}.$$

Also, for every $x \in X$ and for every $n \geq L$, we have

$$d_Z(g(f_n(x)), g_n(f_n(x))) < \tfrac{\varepsilon}{2}.$$

Thus for every $x \in X$ and for every $n \geq L$, we have

$$\begin{aligned} d_Z(g(f(x)), g_n(f_n(x))) &\leq d_Z(g(f(x)), g(f_n(x))) + d_Z(g(f_n(x)), g_n(f_n(x))) \\ &< \tfrac{\varepsilon}{2} + \tfrac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

Therefore $d(g \circ f, g_n \circ f_n) < \varepsilon$ for every $n \geq L$, and so $g_n \circ f_n \to g \circ f$. Since $g_n \circ f_n$ is a homeomorphism for every n, this proves that $g \circ f$ is approximable by homeomorphisms. ☐

Proposition 3.26. *Let M be a compact d-dimensional manifold, possibly with nonempty boundary. Let $\{X_1, \dots, X_n\}$ be a finite collection of pairwise disjoint cellular sets in* $\operatorname{Int} M$*. Then the quotient map $\pi \colon M \to M/\{X_1, \dots, X_n\}$ is approximable by homeomorphisms. In particular, the quotient $M/\{X_1, \dots, X_n\}$ is homeomorphic to M.*

We point out that the space $M/\{X_1, \dots, X_n\}$ is the quotient where the collection of cellular subsets $X_1, \dots, X_n$ is crushed to n distinct points, rather than $M/\bigcup\{X_1, \dots, X_n\}$, where all the X_i are identified to a single point.

Proof of Proposition 3.26. We give a proof by induction. For the case $n = 1$, see Proposition 3.22. Suppose that the quotient map on M crushing any given pairwise disjoint collection of $n - 1$ cellular sets in $\operatorname{Int} M$ is approximable by homeomorphisms, for some $n \geq 2$. The quotient map $\pi \colon M \to M/\{X_1, \dots, X_n\}$ factors as the composition

$$M \to M/X_1 \to M/\{X_1, \dots, X_n\}.$$

The first quotient map is approximable by homeomorphisms by Proposition 3.22. The second map is approximable by homeomorphisms by the inductive hypothesis. Here we are using the fact that $\{X_2, \dots, X_n\}$ is mapped to a pairwise disjoint collection of cellular sets in M/X_1 by the quotient map. This follows from Proposition 3.16 and the fact that the quotient map $M \to M/X_1$ is a homeomorphism when restricted to $M \smallsetminus X_1$. Compositions of maps between compact metric spaces which are approximable by homeomorphisms are themselves approximable by homeomorphisms according to Proposition 3.24. This completes the proof that π is approximable by homeomorphisms. ☐

For the application to the Schoenflies theorem, we will need the following proposition.

Proposition 3.27. *Let $f \colon S^d \to S^d$ be a continuous surjection for some d with exactly two inverse sets A and B. Then each of A and B is cellular.*

To prove Proposition 3.27, we will need the following lemma.

LEMMA 3.28. *Let $f\colon D^d \to S^d$ be a continuous function such that $X \subseteq \operatorname{Int} D^d$ is the only inverse set of f and $f(\operatorname{Int} D^d)$ is open in S^d. Then X is cellular in D^d.*

PROOF. Let D denote the codomain D^d. Let $X = f^{-1}(y)$ for some $y \in S^d$. Since $X \subseteq \operatorname{Int} D$ and we know that $f(\operatorname{Int} D)$ is open in S^d, there is an $\varepsilon > 0$ such that the standard disc $B_\varepsilon(y)$ of radius ε around y in S^d is contained in $f(\operatorname{Int} D)$. Next, choose some $z \in S^d$ not in $\operatorname{Im}(D)$. In particular, we have $z \neq y$. Let V be a standard open ball neighbourhood of z in S^d such that $S^d \smallsetminus V$ is a d-dimensional disc B with $B_\varepsilon(y) \subseteq f(D) \subseteq S^d \smallsetminus V = B$. For every $n \in \mathbb{N}$, choose some homeomorphism $s_{\varepsilon/2^n}\colon S^d \to S^d$ that restricts to the identity on the small disc $B_{\varepsilon/2^{n+1}}(y)$ and squeezes the rest of B into $B_{\varepsilon/2^n}(y)$. In a ball containing B, this could, for instance, be defined radially, and then extended to all of S^d, stretching out V so as to cover the complement of $B_{\varepsilon/2^n}(y)$. Using this function we now define a map $\sigma_{\varepsilon/2^n}\colon D \to D$ by setting

$$\sigma_{\varepsilon/2^n}(x) = \begin{cases} x & \text{if } x \in X \\ f^{-1} \circ s_{\varepsilon/2^n} \circ f(x) & \text{if } x \notin X. \end{cases}$$

Here, $\sigma_{\varepsilon/2^n}$ is defined i.e. f^{-1} may be used, because f is injective on $D \smallsetminus X$ and $s_{\varepsilon/2^n} \circ f$ does not map x to $y = f(X)$ as long as $x \in D \smallsetminus X$. By the closed map lemma (Lemma 3.23), $f\colon D \to S^d$ is a closed map. The restriction $f\colon D \smallsetminus X \to S^d \smallsetminus \{y\}$ is also closed as a restriction of a closed map to a saturated set. As a consequence, the composition

$$f^{-1} \circ s_{\varepsilon/2^n} \circ f|_{D \smallsetminus X}\colon D \smallsetminus X \to D \smallsetminus X$$

is continuous. Define $U := f^{-1}(B_{\varepsilon/2^{n+1}}(y)) \supseteq X$. Then by construction $\sigma_{\varepsilon/2^n}|_U = \operatorname{Id}_U$. We deduce that $\sigma_{\varepsilon/2^n}$ is continuous since both $f^{-1} \circ s_{\varepsilon/2^n} \circ f|_{D \smallsetminus X}$ and $\sigma_{\varepsilon/2^n}|_U$ are continuous. Furthermore, the map $\sigma_{\varepsilon/2^n}$ is injective because the maps $\sigma_{\varepsilon/2^n}|_{D \smallsetminus X}$ and $\sigma_{\varepsilon/2^n}|_X$ are injective and have disjoint images. As a result the image $\operatorname{Im} \sigma_{\varepsilon/2^n} \subseteq D$ is Hausdorff and by the closed map lemma (Lemma 3.23) the map $\sigma_{\varepsilon/2^n}\colon D \to \operatorname{Im} \sigma_{\varepsilon/2^n}$ is a closed map. Thus, the inverse $\sigma^{-1}_{\varepsilon/2^n}\colon \operatorname{Im} \sigma_{\varepsilon/2^n} \to D$ is continuous and $\sigma_{\varepsilon/2^n}$ is an embedding. Therefore $\sigma_{\varepsilon/2^n}(D)$ is homeomorphic to a ball for every n. To finish the proof, observe that the balls $\mathcal{B}_n := \sigma_{\varepsilon/2^n}(D) \subseteq D$, for $n = 1, 2, \ldots$, exhibit X as a cellular set. In particular note that $\mathcal{B}_{n+1} \subseteq \operatorname{Int} \mathcal{B}_n$, since $\sigma_{\varepsilon/2^n}(\partial D)$ lies in $f^{-1}(B_{\varepsilon/2^n})$ but not in $f^{-1}(B_{\varepsilon/2^{n+1}})$. □

PROOF OF PROPOSITION 3.27. In an attempt to reduce confusion, let S and T denote the two copies of S^d. That is, we have a function $f\colon S \to T$. We show that B is cellular. Let $a := f(A)$ and $b := f(B)$. Since A and B are precisely the two inverse sets of f, we know that they are closed and disjoint. Thus, there exists some standard open ball $U \subseteq S$ disjoint from $A \cup B$ such that if $D := S \smallsetminus U$, then D is a standard closed d-dimensional ball and $A \cup B \subseteq \operatorname{Int} D$.

Then we claim that $f(\operatorname{Int} D)$ is open in T. Note that f is a closed map by the closed map lemma (Lemma 3.23). Thus $f(\overline{U})$ is closed. But then $f(\operatorname{Int} D) = T \smallsetminus f(\overline{U})$ is open as claimed.

Then since $a, b \in f(\operatorname{Int} D)$ are distinct, there exists some open set $V \subseteq f(\operatorname{Int} D)$ with $a \in V$ and $b \notin V$. Choose a homeomorphism $h\colon T \to T$ taking $f(D)$ to V bijectively and fixing some smaller neighbourhood $W \subsetneq V$ of a. Recall that $D \subseteq S$. Define a map $\psi\colon D \to S$ as follows.

$$\psi(x) = \begin{cases} x & \text{if } x \in f^{-1}(W) \\ f^{-1} \circ h \circ f(x) & \text{if } x \in D \smallsetminus A. \end{cases}$$

The function above is defined and continuous since f is injective away from A and B and $f^{-1} \circ h \circ f(x) = x$ on $f^{-1}(W)$.

We also check that $B \in \operatorname{Int} D$ is the only inverse set of ψ. This follows since h maps $f(D)$ into V and f is injective away from A and B. To finish, we need to show that $\psi(\operatorname{Int} D)$ is open in S. We check by hand that $\psi(\operatorname{Int} D) = f^{-1} \circ h \circ f(\operatorname{Int} D)$. We know from before that $f(\operatorname{Int} D)$ is open in T. Then $f^{-1} \circ h \circ f(\operatorname{Int} D)$ is open since h is a homeomorphism and f is continuous. Now apply Lemma 3.28 to the map $\psi\colon D \to S$ to conclude that B is cellular. A similar proof shows that A is cellular. □

3.4. Brown's proof of the Schoenflies theorem

After our lengthy interlude in the previous section, we return to the Schoenflies theorem, which we restate below in an equivalent form.

THEOREM 3.29 (Schoenflies theorem [**Bro60**]). *Let $i\colon S^{d-1} \hookrightarrow S^d$ be a continuous embedding admitting a bicollar. Then the closure of each component of $S^d \smallsetminus i(S^{d-1})$ is homeomorphic to D^d.*

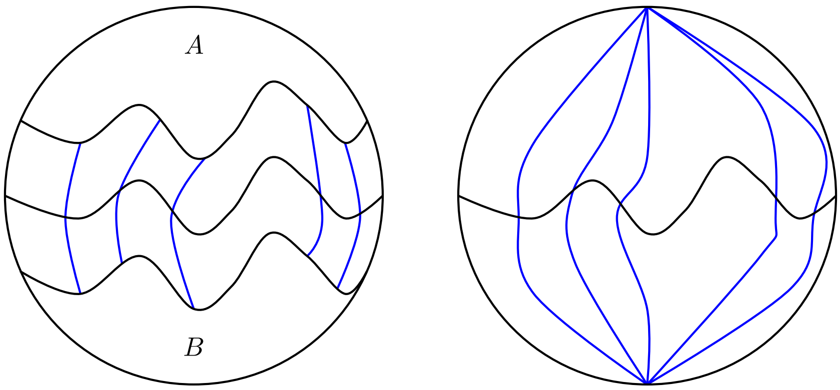


Figure 3.8. Brown's proof of the Schoenflies theorem. Going from left to right, the regions labelled A and B are collapsed to a point each, stretching out a neighbourhood of the equator in the process.

PROOF. By the bicollar hypothesis there exists $J\colon S^{d-1} \times [-1,1] \to S^d$ such that $J|_{S^{d-1}\times\{0\}}$ equals i. From elementary homology computations as in the proof of Theorem 3.5, it follows that the complement of the image of J in S^d has exactly two connected components. Denote their closures by A and B, where A meets $J(S^{d-1} \times \{1\})$.

Observe that the quotient space $S^d/\{A, B\}$ is homeomorphic to S^d, due to the existence of the bicollar. In other words, $S^d/\{A, B\}$ can be identified with the (unreduced) suspension of S^{d-1}, which is homeomorphic to S^d, as indicated in Figure 3.8. Thus we have the composition

$$f\colon S^d \xrightarrow{\pi} S^d/\{A, B\} \xrightarrow{\cong} S^d,$$

where $\cong$ denotes homeomorphism and π denotes the quotient map. This map $f\colon S^d \to S^d$ has exactly two inverse sets, namely A and B, and is surjective. By Proposition 3.27, each of A and B is cellular.

Let D denote the closed northern hemisphere of S^d, thought of as a subset of the codomain S^d. By definition, D is a copy of the d-dimensional ball. Let $U := A \cup (J(S^{d-1} \times (0, 1]))$, that is, U is the component of $S^d \smallsetminus i(S^{d-1})$ containing

A. Then we have the restriction $f|_{\overline{U}}\colon \overline{U} \to D$ whose unique inverse set is A which is cellular in S^d and thus in $\overline{U}$ by Proposition 3.16.

Our goal is to apply Proposition 3.22 to $\overline{U}$. To do so, we first need to show that $\overline{U}$ is a manifold (with boundary). The only possible failure could be near the boundary. As a subspace of S^d, $\overline{U}$ is already Hausdorff and second countable, so we only need to show that it is locally Euclidean. Let $E := J(S^{d-1} \times [0, \frac{1}{2}])$. Note that $\overline{U} = U \cup E$. Moreover f restricts to a continuous bijection from E to some collar of ∂D. This collar of D is closed by the closed map lemma (Lemma 3.23). Therefore $f|_E$ is a homeomorphism, so $\overline{U}$ is a manifold as needed.

Then we have the following diagram:

$$\begin{tikzcd} \overline{U} \arrow[r, "f"] \arrow[d, "\pi"'] & D \\ \overline{U}/A. \arrow[ur, "\overline{f}"'] & \end{tikzcd}$$

The map f is constant on the fibres of the quotient map $\pi\colon \overline{U} \to \overline{U}/A$ and thus descends to the map $\overline{f}$, which is a homeomorphism by the closed map lemma (Lemma 3.23). Next, since A is cellular, the map π is approximable by homeomorphisms by Proposition 3.22. Let $\widetilde{\pi}\colon \overline{U} \to \overline{U}/A$ be any such approximating homeomorphism. Then $\overline{f} \circ \widetilde{\pi}\colon \overline{U} \to D$ is a homeomorphism. It follows that $\overline{U}$ is homeomorphic to the d-dimensional ball $D \cong D^d$ as claimed. $\square$

CHAPTER 4

Decomposition space theory and the Bing shrinking criterion

Christopher W. Davis, Boldizsár Kalmár, Min Hoon Kim, and Henrik Rüping

We begin this chapter with the *Bing shrinking criterion*, characterising the maps between compact metric spaces that are approximable by homeomorphisms. This formalises the shrinking arguments employed in the previous chapter in Brown's proof of the Schoenflies theorem.

Then we study *decompositions* and *decomposition spaces*: a decomposition is a collection of pairwise disjoint subsets of a given space and the corresponding decomposition space is the quotient space arising from identifying each of the constituent subsets of a decomposition to a point. The Bing shrinking criterion is a powerful tool to prove that, in favourable cases, the decomposition space is homeomorphic to the original space, and indeed that the quotient map is approximable by homeomorphisms. We will be interested in studying some fascinating examples of this phenomenon in the chapters to follow. In this chapter, we develop some of the basic theory, showing that the decomposition spaces corresponding to *upper semi-continuous decompositions* of compact metric spaces are well behaved. In particular, the decomposition space of an upper semi-continuous decomposition of a compact metric space is again a compact metric space.

4.1. The Bing shrinking criterion

As demonstrated by Brown's proof of the Schoenflies theorem, shrinking is a powerful tool in the topological category. Here is a natural generalisation of the results from the previous chapter, in the context of shrinking infinitely many sets. The following statement is due to Bing [**Bin52**], and the proof given is from [**Edw80**].

THEOREM 4.1 (Bing shrinking criterion). *Let X and Y be compact metric spaces. Let $f\colon X \to Y$ be a surjective continuous map. Let $W \subseteq X$ be an open set containing every inverse set of f. Then f is approximable by homeomorphisms agreeing with f on $X \setminus W$ if and only if for every $\varepsilon > 0$ there exists a self-homeomorphism $h\colon X \to X$ that restricts to the identity map on $X \setminus W$ and satisfies the following conditions.*

(i) For all $x \in X$, we have that $d_Y\big(f(x), f \circ h(x)\big) < \varepsilon$.
(ii) For all $y \in Y$, we have that $\operatorname{diam}_X h(f^{-1}(y)) < \varepsilon$.

The first condition implies that h is close to the identity as measured in the target space Y and can be restated using the uniform metric as $d(f, f \circ h) < \varepsilon$. The second condition indicates that the inverse sets are shrunk by h. Morally speaking, the theorem says that finding a coherent way of shrinking inverse sets is equivalent to approximating by homeomorphisms.

Proof. Throughout the proof the word 'homeomorphism' will mean a homeomorphism that restricts to the identity map on $X \setminus W$. One direction is relatively easy. Given a sequence of approximating homeomorphisms $h_n\colon X \to Y$ and $\varepsilon > 0$, compositions of the form $h_n^{-1} \circ h_{n+k_n}$ satisfy (i) and (ii) of Theorem 4.1 as long as n and k_n are large enough.

The other, more interesting, direction can be proved by elementary methods. However, we give the proof due to Edwards [Edw80] using the Baire category theorem.

Theorem 4.2 (Baire category theorem). *In a complete metric space, the intersection of any countable collection of open and dense sets is dense.*

Our goal is to construct a uniformly convergent sequence of approximating homeomorphisms for f. Consider the space $\mathcal{C}(X, Y)$ of continuous maps from X to Y equipped with the uniform metric. Since Y is compact, Y is complete, and therefore the space $\mathcal{C}(X, Y)$ is a complete metric space [Mun00, Theorem 43.6]. Let E denote the closure of the set

$$S := \{f \circ h \mid h\colon X \to X \text{ is a homeomorphism}\}.$$

Note that $f \in E$, by taking $h = \mathrm{Id}$. Moreover, $E \subseteq \mathcal{C}(X, Y)$ is a closed subset of a complete metric space and is thus itself a complete metric space. For each positive integer n, let $E_{\frac{1}{n}}$ denote the set

$$E_{\frac{1}{n}} := \{g \in E \mid \mathrm{diam}_X\, g^{-1}(y) < \tfrac{1}{n} \text{ for every } y \in Y\}.$$

By hypothesis, each $E_{\frac{1}{n}}$ is nonempty. More precisely, take some $\varepsilon < \frac{1}{n}$ and find an h satisfying (ii) in the hypotheses. Then $g = f \circ h^{-1} \in E_{\frac{1}{n}}$, so $E_{\frac{1}{n}}$ is nonempty.

Next we show that $E_{\frac{1}{n}}$ is open in E. This is the content of the next two claims.

Claim. *For each $g \in E_{\frac{1}{n}}$ there is an $\alpha > 0$ such that whenever $d_X(x, x') \geq 1/n$, then $d_Y(g(x), g(x')) > \alpha$.*

To see the claim, suppose for a contradiction that it is false. Then there would exist sequences $\{x_i\}$, $\{x'_i\}$ in X such that $d_X(x_i, x'_i) \geq \frac{1}{n}$ for all i, but $d_Y(g(x_i), g(x'_i))$ converges to 0. Since $X \times X$ is a compact metric space, there is a convergent subsequence of $\{(x_i, x'_i)\}$ with limit $(x, x') \in X \times X$. Note that $d_X(x, x') \geq \frac{1}{n}$.

We assert that $g(x) = g(x')$. Suppose that this were false; then write $\beta := d_Y(g(x), g(x')) > 0$. Choose an N such that

$$d_Y(g(x), g(x_i)) < \beta/3$$

and

$$d_Y(g(x'), g(x'_i)) < \beta/3$$

whenever $i \geq N$. We have then, for every $i \geq N$, that

$$\begin{aligned}\beta = d_Y(g(x), g(x')) &\leq d_Y(g(x), g(x_i)) + d_Y(g(x_i), g(x'_i)) + d_Y(g(x'_i), g(x')) \\ &< 2\beta/3 + d_Y(g(x_i), g(x'_i)).\end{aligned}$$

Thus $\beta/3 < d_Y(g(x_i), g(x'_i))$ for all $i \geq N$, which contradicts the hypothesis that $d_Y(g(x_i), g(x'_i)) \to 0$ as $i \to \infty$. Thus $g(x) = g(x')$ as asserted.

But $g(x) = g(x')$ implies that x and x' both lie in $g^{-1}(g(x))$. Then with $y = g(x)$ we know that $\mathrm{diam}_X\, g^{-1}(y) < \frac{1}{n}$ since $g \in E_{\frac{1}{n}}$. Thus, we have that $d_X(x, x') < \frac{1}{n}$. This contradicts the hypothesis that $d_X(x, x') \geq \frac{1}{n}$, which completes the proof of the claim. We use the claim just proven in the proof of the next claim.

Claim. *Let $g \in E_{\frac{1}{n}}$. For α as in the previous claim, if $g' \in E$ satisfies $d(g, g') < \frac{\alpha}{2}$, then $g' \in E_{\frac{1}{n}}$.*

Suppose that the claim is false. Let $g' \in E$ such that $d(g, g') < \frac{\alpha}{2}$ but $g' \notin E_{\frac{1}{n}}$. Then there exist $x, x' \in X$ with $d_X(x, x') \geq \frac{1}{n}$ but $g'(x) = g'(x') = y$ for some $y \in Y$. Then

$$\begin{aligned} d_Y(g(x), g(x')) &\leq d_Y(g(x), y) + d_Y(g(x'), y) = d_Y(g(x), g'(x)) + d_Y(g(x'), g'(x')) \\ &< \alpha/2 + \alpha/2 = \alpha, \end{aligned}$$

where for the last inequality we are using that $d(g, g') < \frac{\alpha}{2}$. However, by the previous claim, $d_Y(g(x), g(x')) > \alpha$, which is a contradiction, proving the current claim.

We have now shown that $E_{\frac{1}{n}}$ is open as desired, since for any fixed $g \in E_{\frac{1}{n}}$, we have shown that there exists $\alpha > 0$ so that every g' with $d(g, g') < \frac{\alpha}{2}$ is contained in $E_{\frac{1}{n}}$. In other words, we have found an open neighbourhood of g in $E_{\frac{1}{n}}$.

Now we show that $E_{\frac{1}{n}}$ is dense in E for each n. Let $g \in E$, and let $N \subseteq E$ be a neighbourhood of g in E. Let $\eta > 0$ be such that $B_\eta(g) \subseteq N$. To show that $E_{\frac{1}{n}}$ is dense, we want to show that there is a $\widetilde{g} \in B_\eta(g) \cap E_{\frac{1}{n}}$. This will show that every neighbourhood of g intersects $E_{\frac{1}{n}}$.

Since E is the closure of S, there is a homeomorphism $h\colon X \to X$ such that $d(f \circ h, g) < \frac{\eta}{2}$. Use conditions (i) and (ii) in the hypotheses of Theorem 4.1 to obtain a homeomorphism $H\colon X \to X$ for which

(a) $d(f, f \circ H) < \frac{\eta}{2}$, and
(b) for every $y \in Y$ the diameter $\operatorname{diam}_X H(f^{-1}(y))$ is small enough so that

$$\operatorname{diam}_X h^{-1}(H(f^{-1}(y))) < \tfrac{1}{n},$$

as follows. Note that $h^{-1}\colon X \to X$ is uniformly continuous by the Heine-Cantor theorem, since X is compact. Thus there exists $\delta > 0$ such that $d_X(h^{-1}(x), h^{-1}(x')) < \frac{1}{n}$ whenever $d_X(x, x') < \delta$. We then have H from the hypotheses of Theorem 4.1 using $\varepsilon = \min\{\eta/2, \delta\}$. As a result $\operatorname{diam}_X H(f^{-1}(y)) < \delta$, from which it follows that $\operatorname{diam}_X h^{-1}(H(f^{-1}(y))) < \frac{1}{n}$ as desired, as well as $d(f, f \circ H) < \frac{\eta}{2}$.

Denote the map $f \circ H^{-1} \circ h$ by $\widetilde{g}$. Then $\operatorname{diam}_X \widetilde{g}^{-1}(y) < \frac{1}{n}$ for every $y \in Y$ and thus $\widetilde{g} \in E_{\frac{1}{n}}$. In addition, we have

$$d(g, \widetilde{g}) \leq d(g, f \circ h) + d(f \circ h, \widetilde{g}) < \tfrac{\eta}{2} + d(f \circ h, \widetilde{g}).$$

But

$$\begin{aligned} d(f \circ h, \widetilde{g}) &= d(f \circ h, f \circ H^{-1} \circ h) = d(f, f \circ H^{-1}) = d(f \circ H, f \circ H^{-1} \circ H) \\ &= d(f \circ H, f) < \tfrac{\eta}{2}, \end{aligned}$$

where the first equality follows from the definition of $\widetilde{g}$ and the second and third from the fact that h and H are bijections. Thus, we see that $d(g, \widetilde{g}) < \frac{\eta}{2} + \frac{\eta}{2} = \eta$, so $\widetilde{g} \in B_\eta(g)$. This means that every neighbourhood N of g intersects $E_{\frac{1}{n}}$, and so $E_{\frac{1}{n}}$ is dense in E as desired.

Now, by the Baire category theorem (Theorem 4.2), the set $E_0 = \bigcap_n E_{\frac{1}{n}}$ is dense in E. In particular, E_0 is nonempty and contains maps arbitrarily close to f. But every element of E_0 is a bijection since the inverse sets have to be points. By the closed map lemma, every continuous map from a compact space to a Hausdorff space is closed, and thus the elements of E_0 are homeomorphisms. $\square$

4.2. Decompositions

The Bing shrinking criterion (Theorem 4.1) leads us to the field of decomposition space theory in general. As in the previous chapter, we will be interested in crushing pairwise disjoint subsets to (distinct) points. However, unlike in the previous chapter, we will usually be interested in crushing infinitely many subsets, possibly even uncountably many subsets, of a given space to points. We will investigate whether the resulting object is homeomorphic to the original space. When considering infinitely many subsets, it might not be possible to shrink everything in a controlled manner as in Proposition 3.26, since whenever we shrink some subsets, the others might stretch out, leading to a subtle and beautiful theory. We will investigate these questions in depth for some interesting examples in the next few chapters.

In the rest of this chapter we cover some of the foundations of the field of decomposition space theory. An extensive account is given in [**Dav07**], although the terminology used therein differs slightly from ours.

Definition 4.3. A *decomposition* of a topological space X is a collection

$$\mathcal{D} = \{\Delta_i\}_{i \in I}$$

of pairwise disjoint subsets $\Delta_i \subseteq X$, called the *decomposition elements*, indexed by some index set I.

Given a decomposition $\mathcal{D}$ of X, the *decomposition space* is the quotient space $X/\mathcal{D}$ obtained by factoring out X by the sets $\Delta \in \mathcal{D}$. Colloquially, each Δ_i is crushed to a point. We endow $X/\mathcal{D}$ with the quotient topology and usually denote the corresponding quotient map by $\pi \colon X \to X/\mathcal{D}$.

Note that $X/\mathcal{D}$ is different from $X/\bigcup_{i \in I} \Delta_i$ where the entire set $\bigcup_{i \in I} \Delta_i$ is crushed to a single point.

Remark 4.4. In the literature, a decomposition is often required to be a partition of the given space. Of course, any decomposition $\mathcal{D} = \{\Delta_i\}_{i \in I}$ for a space X as in Definition 4.3 can be completed to a partition by adding singletons for each point in $X \setminus \bigcup_{i \in I} \Delta_i$ without altering the corresponding decomposition space. We prefer our definition since it reduces the number of sets that need to be specified. On the other hand, it is permitted, and may often happen, that some decomposition elements *are* singleton sets.

In practice, when we wish to describe a decomposition, we will usually only describe a *defining sequence*.

Definition 4.5. A *defining sequence* for a decomposition of a topological space X is a sequence $\{C_i\}_{i=1}^{\infty}$ where each $C_i \subseteq X$ is compact and $C_{i+1} \subseteq \operatorname{Int} C_i$ for each i. The decomposition elements of the decomposition associated with this defining sequence are the connected components of $\bigcap_{i=1}^{\infty} C_i$.

The defining sequence for a decomposition will often be produced by an iterated construction using a single pattern, which we call a *defining pattern*. We explain this process by the following example.

Example 4.6. We give the defining sequence and a corresponding defining pattern for the decomposition that we will study in detail in the next chapter. On the left of Figure 4.1 we show a 2-component link embedded in the standard unknotted solid torus in S^3. This link is the *Bing double* of the core of the solid torus, and we have already encountered it in Chapter 1. Thicken up the two components of the link to solid tori $S^1 \times D^2$ embedded in the interior of the larger $S^1 \times D^2$. This is the defining pattern.

Now iterate the embedding: at each stage, embed the same pattern, with no twisting, inside each of the solid tori of the previous stage. Note that this doubles the number of components. On the right of Figure 4.1, we show the second stage of this process. Each iteration produces the next term of the defining sequence. The corresponding decomposition, which is called the *Bing decomposition*, consists of

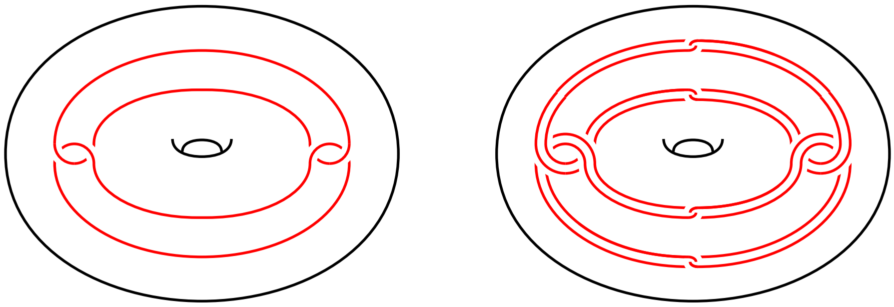


Figure 4.1. The defining pattern of the Bing decomposition (left) and its second stage (right).

the connected components of the infinite intersection of these nested solid tori, by definition. We shall return to this example in detail in Chapter 5.

In our discussion of decompositions we will often wish to modify the defining sequence of a decomposition by isotopies. The following proposition shows that this does not change the homeomorphism type of the corresponding decomposition space.

Proposition 4.7. *Let $\{C_i\}_{i=1}^{\infty}$ be a defining sequence for a decomposition $\mathcal{D}$ for a d-dimensional manifold M, where each C_i is a d-dimensional submanifold with boundary within M. Let $\phi_0\colon M \to M$ be a homeomorphism. For each $i \geq 1$, let $\phi_i\colon C_i \to C_i$ be a homeomorphism restricting to the identity on the boundary. Extend ϕ_i over M via the identity on $M \smallsetminus C_i$. Let $\mathcal{D}'$ denote the decomposition of M given by the defining sequence $C_i' = \phi_{i-1} \circ \cdots \circ \phi_0(C_i)$ for each i. Then the decomposition spaces $M/\mathcal{D}$ and $M/\mathcal{D}'$ are homeomorphic.*

Proof. Construct a homeomorphism between $M/\mathcal{D}$ and $M/\mathcal{D}'$ using the map ϕ_0 on $M \smallsetminus C_1$ and the map $\phi_i \circ \cdots \circ \phi_0$ on $C_i \smallsetminus C_{i+1}$, and by sending each component of $\bigcap_{i=1}^{\infty} C_i$ (that is, any decomposition element) to the corresponding component of $\bigcap_{i=1}^{\infty} C_i'$. The infinite composition is defined since any given point is nontrivially affected by at most finitely many ϕ_i. $\square$

In light of the above proposition, we will regard two decompositions as the same if they can be related by such a sequence of homeomorphisms. We will see in Section 5.2 that the homeomorphism type of the individual decomposition elements need not be preserved by this operation. However, since we are only interested in the homeomorphism type of the entire decomposition space, it is only ever necessary (for us) to consider decompositions up to this notion of equivalence.

4.3. Upper semi-continuous decompositions

This section contains some of the elementary theory of decomposition spaces. We will concern ourselves with *upper semi-continuous decompositions*, which we soon define. We will show in this section that the decomposition spaces for these decompositions are particularly well behaved.

Definition 4.8. Let X be a space with a decomposition $\mathcal{D} = \{\Delta_i\}_{i \in I}$ and let $\pi\colon X \to X/\mathcal{D}$ denote the quotient map. Given a subset $S \subseteq X$, define its *$\mathcal{D}$-saturation* as

$$\pi^{-1}(\pi(S)) = S \cup \bigcup\{\Delta_i \mid \Delta_i \cap S \neq \emptyset\}.$$

We say that S is $\mathcal{D}$*-saturated* if $S = \pi^{-1}(\pi(S))$, that is if S is saturated with respect to π. Note that the $\mathcal{D}$-saturation of S is the smallest saturated subset of X (with respect to $\pi\colon X \to X/\mathcal{D}$) that contains S.

For a subset $S \subseteq X$, we also consider the largest saturated subset of S, defined by

$$S^* := S \smallsetminus \pi^{-1}\big(\pi(X \smallsetminus S)\big) = S \smallsetminus \bigcup\{\Delta_i \mid \Delta_i \not\subseteq S\}.$$

Definition 4.9. A decomposition $\mathcal{D}$ of a topological space X is said to be *upper semi-continuous* if for each open subset $U \subseteq X$, the set U^* is also open and each decomposition element $\Delta \in \mathcal{D}$ is closed and compact.

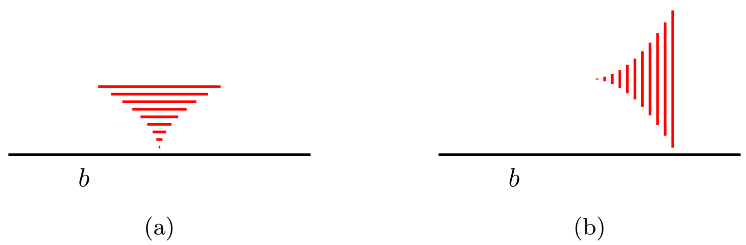


Figure 4.2. Two decompositions of the plane are shown. Each line segment depicted above is closed. In each case, there is precisely one black decomposition element, denoted by b, and infinitely many red decomposition elements. On the left, the red decomposition elements decrease in length converging to b. On the right, the lengths of the red decomposition elements strictly increase as we move to the right and the bottom end points converge to b.
The decomposition on the left is easily seen to be upper semi-continuous by Proposition 4.10(3), since we can find small saturated neighbourhoods for each decomposition element. For the decomposition on the right, the black decomposition element b does not have small saturated neighbourhoods and thus the decomposition is not upper semi-continuous, by Proposition 4.10(3).

Now we give some useful equivalent formulations for upper semi-continuity of decompositions. For example, we can use these characterisations to check that Figure 4.2 gives an example, and a non-example, of an upper semi-continuous decomposition of the plane.

Proposition 4.10. *Let $\mathcal{D} = \{\Delta_i\}_{i\in I}$ be a decomposition of a space X with Δ_i closed and compact for every i. Then the following are equivalent.*

(1) The decomposition $\mathcal{D}$ is upper semi-continuous.
(2) The quotient map $\pi\colon X \to X/\mathcal{D}$ is closed.
(3) If U is an open neighbourhood of C, where C is either some Δ_i or $\{x\}$ for some $x \in X \smallsetminus \bigcup_{i\in I} \Delta_i$, then U contains a $\mathcal{D}$-saturated open neighbourhood of C.

Proof. (1) $\Rightarrow$ (2) Suppose that $\mathcal{D}$ is an upper semi-continuous decomposition. For every closed subset $C \subseteq X$, the set $(X \smallsetminus C)^*$ is open by definition. Note that $(X \smallsetminus C)^* = X \smallsetminus \pi^{-1}(\pi(C))$, and thus $\pi^{-1}(\pi(C))$ is closed in X. Hence $\pi(C)$ is closed in $X/\mathcal{D}$ for every closed $C \subseteq X$ by definition of the quotient topology.

(2) $\Rightarrow$ (1) Suppose that $\pi\colon X \to X/\mathcal{D}$ is a closed map. Then for any open subset $U \subseteq X$, $U^* = U \smallsetminus \pi^{-1}(\pi(X \smallsetminus U))$ is open since $X \smallsetminus U$ and π are closed, and π is continuous. Each Δ_i is closed and compact by hypothesis.

(1) $\Rightarrow$ (3) Given any such set U, the set U^* is a $\mathcal{D}$-saturated neighbourhood of C.

(3) $\Rightarrow$ (1) Given an open set $U \subseteq X$, let $x \in U^*$. Then, either there exists Δ_i such that $x \in \Delta_i \subseteq U^*$, or $x \in U^* \smallsetminus \bigcup_{i \in I} \Delta_i$ since U^* is $\mathcal{D}$-saturated. In either case, by hypothesis, there is some $\mathcal{D}$-saturated open neighbourhood of x within U. Since this neighbourhood is $\mathcal{D}$-saturated, it lies in U^* as needed. Thus U^* is open. Each Δ_i is closed and compact by hypothesis. $\square$

We record the following property of upper semi-continuous decompositions for later use in Remark 7.6.

Proposition 4.11. *Suppose that $\mathcal{D}_i$ are upper semi-continuous decompositions of the spaces X_i for $i = 1, 2$. Then the space $(X_1/\mathcal{D}_1) \times (X_2/\mathcal{D}_2)$ is homeomorphic to the space $(X_1 \times X_2)/(\mathcal{D}_1 \times \mathcal{D}_2)$, where the decomposition $\mathcal{D}_1 \times \mathcal{D}_2$ of $X_1 \times X_2$ has decomposition elements of the form $\Delta_1^i \times \Delta_2^i$, $\Delta_1^i \times \{x_2\}$, or $\{x_1\} \times \Delta_2^i$, where $\Delta_1^i \in \mathcal{D}_1$; $\Delta_2^i \in \mathcal{D}_2$; and $x_i \in X_i \smallsetminus \bigcup \mathcal{D}_i$ for all i.*

Proof. Define $f\colon (X_1 \times X_2)/(\mathcal{D}_1 \times \mathcal{D}_2) \to (X_1/\mathcal{D}_1) \times (X_2/\mathcal{D}_2)$ by $[(x_1, x_2)] \mapsto ([x_1], [x_2])$. As a set map, f is well defined and bijective by construction. The following diagram commutes, where p is the quotient map $X_1 \times X_2 \to (X_1 \times X_2)/(\mathcal{D}_1 \times \mathcal{D}_2)$ and $\pi = \pi_1 \times \pi_2$ is the product of the quotient maps $\pi_i\colon X_i \to X_i/\mathcal{D}_i$.

$$\begin{tikzcd}
X_1 \times X_2 \arrow[r, "\pi"] \arrow[d, "p"'] & (X_1/\mathcal{D}_1) \times (X_2/\mathcal{D}_2) \\
(X_1 \times X_2)/(\mathcal{D}_1 \times \mathcal{D}_2) \arrow[ur, "f"'] &
\end{tikzcd}$$

Since $\pi = f \circ p$ is continuous, the defining property of quotient spaces implies that f is continuous. By Proposition 4.10, π_1 and π_2 are closed maps. Since $\mathcal{D}_i$ is upper semi-continuous, $\pi_i^{-1}(y)$ is compact for every $y \in X_i/\mathcal{D}_i$ and for each i. It follows that the product $\pi = \pi_1 \times \pi_2$ is closed (see for example [**Him65**, Theorem 1]). We remark that compactness of $\pi_i^{-1}(y)$ for each i, y is required since the product of two closed maps is not closed in general. If $C \subseteq (X_1 \times X_2)/(\mathcal{D}_1 \times \mathcal{D}_2)$ is closed, then $p^{-1}(C)$ is closed and hence $f(C) = \pi(p^{-1}(C))$ is closed. We see that the bijective map f is continuous and closed, and so is a homeomorphism. $\square$

We are ready to prove the key property of upper semi-continuous decompositions, namely that a decomposition space for a compact metric space with respect to an upper semi-continuous decomposition is itself a compact metric space.

Proposition 4.12. *Let $\mathcal{D}$ be an upper semi-continuous decomposition of a topological space X.*

(1) *If X is Hausdorff, then $X/\mathcal{D}$ is also Hausdorff.*
(2) *If X is second countable, then $X/\mathcal{D}$ is also second countable.*

Proof. Let $\pi\colon X \to X/\mathcal{D}$ be the quotient map. Since $\mathcal{D}$ is upper semi-continuous, by Proposition 4.10, we know that π is a closed surjective map. Moreover, each $\pi^{-1}(y)$ is compact for $y \in X/\mathcal{D}$, since it is either a decomposition element or a singleton set.

(1) Let y_1 and y_2 be two distinct points in $X/\mathcal{D}$. Then $\pi^{-1}(y_1)$ and $\pi^{-1}(y_2)$ are disjoint compact subsets in X. Since each $\pi^{-1}(y_i)$ is compact and X is Hausdorff, it is elementary to show that there are two disjoint open subsets U_1 and U_2 in X such that $\pi^{-1}(y_i) \subseteq U_i$. Since π is a closed map, each set $\pi(X \smallsetminus U_i)$ is closed, and thus we have open neighbourhoods

$$W_i := X/\mathcal{D} \smallsetminus \pi(X \smallsetminus U_i)$$

of y_i. Note that if $x \in X \smallsetminus U_i$, then $\pi(x) \notin W_i$. That is, $\pi^{-1}(W_i) \subseteq U_i$. Since U_1 and U_2 are disjoint, W_1 and W_2 are also disjoint. This shows that $X/\mathcal{D}$ is Hausdorff.

(2) Let $\{U_i\}_{i \in \mathbb{N}}$ be a countable basis for the topology on X. For each finite subset $I \subseteq \mathbb{N}$, let

$$W_I := X/\mathcal{D} \smallsetminus \pi(X \smallsetminus \bigcup_{i \in I} U_i).$$

Since π is a closed map,

$$\{W_I \mid I \subseteq \mathbb{N} \text{ is finite}\}$$

is a countable collection of open subsets of $X/\mathcal{D}$. We next show that $\{W_I \mid I \subseteq \mathbb{N} \text{ is finite}\}$ is a basis for the topology on $X/\mathcal{D}$.

Let y be a point in $X/\mathcal{D}$ and let W be an open neighbourhood of y. Express $\pi^{-1}(W)$ as a union of elements of $\{U_i\}_{i \in \mathbb{N}}$. Then since $\pi^{-1}(y)$ is a compact subset of $\pi^{-1}(W)$, there is a finite subset $I \subseteq \mathbb{N}$ such that $\pi^{-1}(y) \subseteq \bigcup_{i \in I} U_i \subseteq \pi^{-1}(W)$. Then $y \in W_I = X/\mathcal{D} \smallsetminus \pi(X \smallsetminus \bigcup_{i \in I} U_i)$, since otherwise $\pi^{-1}(y) \cap X \smallsetminus \bigcup U_i$ would be nonempty. Moreover, we have

$$X/\mathcal{D} \smallsetminus W = \pi(X \smallsetminus \pi^{-1}(W)) \subseteq \pi(X \smallsetminus \bigcup_{i \in I} U_i) = X/\mathcal{D} \smallsetminus W_I.$$

This implies that $W_I \subseteq W$. Therefore, $X/\mathcal{D}$ is second countable with a countable basis $\{W_I \mid I \subseteq \mathbb{N} \text{ finite}\}$. $\square$

COROLLARY 4.13. *Let $\mathcal{D}$ be an upper semi-continuous decomposition of a compact metric space X. Then $X/\mathcal{D}$ is a compact metric space.*

PROOF. Since X is a compact metric space, it is Hausdorff and second countable. Compact, Hausdorff spaces are normal [**Mun00**, Theorem 32.3] and normal spaces are regular, so X is regular. Since $\mathcal{D}$ is upper semi-continuous, $X/\mathcal{D}$ is Hausdorff and second countable by Proposition 4.12. Compactness of $X/\mathcal{D}$ is immediate since X is compact. Since $X/\mathcal{D}$ is regular and second countable, by Urysohn's metrisation theorem (see e.g. [**Mun00**, Theorem 34.1]), $X/\mathcal{D}$ is metrisable. See also [**Wil70**, Corollary 23.2, p. 166] which shows that the continuous image of a compact metric space in a Hausdorff space is metrisable. $\square$

REMARK 4.14. While Corollary 4.13 is stated only for compact metric spaces, in fact the decomposition space of any metric space with respect to an upper semi-continuous decomposition is metrisable [**Sto56**, **Han54**, **McA54**] (see also [**Dav07**, Proposition 2.2]).

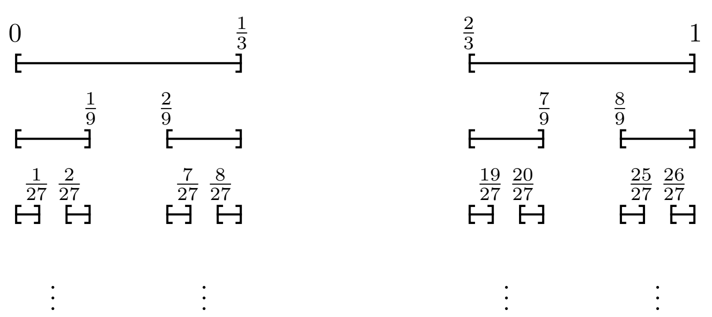


Figure 4.3. Constructing the ternary Cantor set.

Remark 4.15. Corollary 4.13 asserts the existence of a metric on a decomposition space. This abstractly available metric may be elusive in practice. An illuminating 1-dimensional example of this phenomenon comes from the ternary Cantor set $\mathfrak{C}$. Consider the decomposition of the unit interval given by the closures of the 'middle third' intervals which are removed in the construction of $\mathfrak{C}$ (see Figure 4.3). In other words, consider the decomposition of $[0,1]$ given by the connected components of $[0,1] \smallsetminus \mathfrak{C}$. From Proposition 4.10(3), it is easy to see that the decomposition is upper semi-continuous, and thus Corollary 4.13 says that the decomposition space is metrisable. Indeed, the decomposition space is homeomorphic to the interval, which later we will see as an easy corollary of Theorem 9.16 on the shrinkability of null decompositions with recursively starlike-equivalent elements (see Example 9.17).

How could one measure length in this decomposition space? A natural attempt consists of imagining a taxi driving through the interval and turning the meter off when it is within any decomposition element. Unfortunately, since the Cantor set has measure zero, the taxi ride would be free, and so this does not produce a metric. Despite the failure of this naïve attempt, we do know by Corollary 4.13 that some abstract metric exists on the decomposition space. One needs such a metric in order to be able to apply the Bing shrinking criterion (Theorem 4.1). For the case at hand, once we know the decomposition shrinks, and therefore that the decomposition space is homeomorphic to $[0,1]$, of course there is no problem defining a metric, by pulling back the standard metric from $[0,1]$. But this pull back metric cannot be the "taxi metric". In general, we may not be able to produce, even *a fortiori*, a nice metric that we can attempt to visualise. But it will suffice for our purposes to know that a metric exists.

4.4. Shrinkability of decompositions

Now that we have singled out the appropriate class of spaces and decompositions for our purposes, namely compact metric spaces and upper semi-continuous decompositions, we introduce the important concept of *shrinkability* for decompositions. With the Bing shrinking criterion (Theorem 4.1) and the aim of deciding when a quotient map $\pi\colon X \to X/\mathcal{D}$ is approximable by homeomorphisms in mind, we make the following definition.

Definition 4.16 (Shrinkability). Let $\mathcal{D}$ be an upper semi-continuous decomposition of a compact metric space X. The decomposition $\mathcal{D}$ is *shrinkable* if for every $\varepsilon > 0$ there exists a self-homeomorphism $h\colon X \to X$ such that

(i) for all $x \in X$, we have that $d_{X/\mathcal{D}}\big(\pi(x), \pi \circ h(x)\big) < \varepsilon$, and
(ii) for all $\Delta \in \mathcal{D}$, we have $\operatorname{diam}_X h(\Delta) < \varepsilon$,

where $d_{X/\mathcal{D}}$ is some chosen metric on $X/\mathcal{D}$.

Given an open set $W \subseteq X$ containing all the non-singleton decomposition elements, we say $\mathcal{D}$ is *shrinkable fixing* $X \smallsetminus W$ if for every $\varepsilon > 0$ there exists a homeomorphism $h\colon X \to X$ with properties (i) and (ii), and that fixes the set $X \smallsetminus W$ pointwise.

We say the decomposition $\mathcal{D}$ is *strongly shrinkable* if for every open set $W \subseteq X$ containing all the non-singleton decomposition elements, $\mathcal{D}$ is shrinkable fixing $X \smallsetminus W$.

Remark 4.17. For manifolds, shrinkable and strongly shrinkable are equivalent notions [**Dav07**, p. 108, Theorem 13.1]. A counterexample to the general statement for non-manifolds can be constructed using [**Dav07**, Examples 7.1 and 7.2], as indicated in [**Dav07**, p. 112, Exercise 13.2].

Note that the above definition merely applies the Bing shrinking criterion to the quotient map $\pi\colon X \to X/\mathcal{D}$. In other words, a decomposition is shrinkable if and only if the corresponding quotient map $\pi\colon X \to X/\mathcal{D}$ is approximable by homeomorphisms. This is formalised by the following theorem.

THEOREM 4.18. *Let $\mathcal{D}$ be an upper semi-continuous decomposition of a compact metric space X. The decomposition $\mathcal{D}$ is shrinkable if and only if the quotient map $\pi\colon X \to X/\mathcal{D}$ is approximable by homeomorphisms.*

Moreover, given an open set $W \subseteq X$ containing all the non-singleton elements of $\mathcal{D}$, the decomposition $\mathcal{D}$ is shrinkable fixing $X \setminus W$ if and only if π is approximable by homeomorphisms such that each of the homeomorphisms agrees with π on $X \setminus W$.

The decomposition $\mathcal{D}$ is strongly shrinkable if and only if π is approximable by homeomorphisms such that each of the homeomorphisms agrees with π on $X \setminus W$ for every open set $W \subseteq X$ containing all the non-singleton decomposition elements.

Thus in order to show that the quotient map $\pi\colon X \to X/\mathcal{D}$ is approximable by homeomorphisms, or equivalently, that the decomposition $\mathcal{D}$ of X shrinks, it suffices to construct a family of self-homeomorphisms $h\colon X \to X$ having properties (i) and (ii) in Definition 4.16. In the next few chapters we investigate a series of interesting decomposition spaces. In each case, we will either construct such self-homeomorphisms, or show that they cannot exist.

REMARK 4.19. As obliquely indicated by Definition 4.16, the choice of metric on the decomposition space $X/\mathcal{D}$ for a compact metric space X with respect to an upper semi-continuous decomposition $\mathcal{D}$ does not affect the shrinkability of $\mathcal{D}$. One way to see this is via Theorem 4.18. Given metrics d_1 and d_2 on $X/\mathcal{D}$ inducing its topology, the identity map $\mathrm{Id}\colon (X/\mathcal{D}, d_1) \to (X/\mathcal{D}, d_2)$ is uniformly continuous by the Heine-Cantor theorem. Then by composing with this identity map it is easy to see that the quotient map from X to $(X/\mathcal{D}, d_1)$ is approximable by homeomorphisms if and only if the quotient map from X to $(X/\mathcal{D}, d_2)$ is approximable by homeomorphisms.

We finish the chapter by considering cellular sets.

DEFINITION 4.20. A decomposition $\mathcal{D}$ of a manifold M is said to be *cellular* if each element of $\mathcal{D}$ is cellular.

In Chapter 3, we saw some examples of (finite) cellular decompositions which shrink. We now show that every decomposition of a manifold that shrinks must be cellular. However, we will soon see that there exist cellular decompositions of manifolds that do not shrink.

PROPOSITION 4.21. *Let $\mathcal{D}$ be an upper semi-continuous decomposition of a compact n-dimensional manifold M. If $\mathcal{D}$ shrinks, then $\mathcal{D}$ is cellular.*

PROOF. Fix a metric on M. Since M is a compact metric space and $\mathcal{D}$ is upper semi-continuous, the decomposition space $M/\mathcal{D}$ is a compact metrisable space by Corollary 4.13. Fix a metric on $M/\mathcal{D}$. Let $\pi\colon M \to M/\mathcal{D}$ be the quotient map. Let $\Delta \in \mathcal{D}$ and let $W \subseteq M$ be an open set with $\Delta \subseteq W$. It will suffice to find an open set B with closure $\overline{B} \cong D^n$ and with $\Delta \subseteq B \subseteq \overline{B} \subseteq W$.

Let $4\delta > 0$ be the distance from $\pi(\Delta)$ to the compact set $\pi(M \setminus W)$, where the distance is positive since $\pi(\Delta) \notin \pi(M \setminus W)$. For $i = 1, 2, 3$, define $U_i \subseteq M/\mathcal{D}$ to be the open ball of radius $i\delta$ centred at $\pi(\Delta)$. Observe that the collection

$$\mathcal{U} := \{M \setminus \pi^{-1}(\overline{U}_2), \pi^{-1}(U_3 \setminus \overline{U}_1), \pi^{-1}(U_2) \setminus \Delta, \pi^{-1}(U_1)\}$$

forms an open cover of M. Let $\mathcal{V}$ be an open cover of M such that for all $V \in \mathcal{V}$, we have that $\overline{V} \cong D^n$ and moreover, the closure $\overline{V}$ is contained in some element of $\mathcal{U}$. Note that this last condition is stronger than $\mathcal{V}$ being a refinement of $\mathcal{U}$, but is nonetheless satisfied, since M is an n-manifold.

Since M is compact, by the Lebesgue number lemma (see e.g. [**Mun00**, Lemma 27.5]) there exists a Lebesgue number $\eta \in \mathbb{R}$ for the cover $\mathcal{V}$, such that every set of diameter less than η is contained in an element of $\mathcal{V}$. Let $\varepsilon := \min\{\delta, \eta\}$.

Since the decomposition $\mathcal{D}$ shrinks, by definition (Definition 4.16), there exists a homeomorphism $h\colon M \to M$ such that $d_{M/\mathcal{D}}(\pi(x), \pi \circ h(x)) < \varepsilon$ for all $x \in M$ and $\operatorname{diam}_M h(\Delta) < \varepsilon$ for all $\Delta \in \mathcal{D}$. Then since $\operatorname{diam}_M h(\Delta) < \varepsilon \leq \eta$, there is some $B' \in \mathcal{V}$ with $h(\Delta) \subseteq B'$ and $\overline{B'} \cong D^n$. Define $B := h^{-1}(B')$. Since h is a homeomorphism, we have that $\overline{B} \cong D^n$.

Now we show that $\Delta \subseteq B \subseteq \overline{B} \subseteq W$, which will complete the proof. We need only show that $\overline{B} \subseteq W$ since $\Delta \subseteq B$ by construction. Suppose for a contradiction that $\overline{B} \not\subseteq W$. Then let $w \in \overline{B} \smallsetminus W \subseteq M \smallsetminus W$. Then we claim that $h(w) \in M \smallsetminus \pi^{-1}(\overline{U}_2)$. To see this, suppose for a contradiction (within a contradiction) that $h(w) \in \pi^{-1}(\overline{U}_2)$. Then $\pi \circ h(w) \in \overline{U}_2$. Moreover, since $w \in M \smallsetminus W$, we know that $\pi(w) \in \pi(M \smallsetminus W)$. By hypothesis, $d_{M/\mathcal{D}}(\pi(w), \pi \circ h(w)) < \varepsilon$. Then we have

$$\begin{aligned}
4\delta = d_{M/\mathcal{D}}(\pi(\Delta), \pi(M \smallsetminus W)) &\leq d_{M/\mathcal{D}}(\pi(\Delta), \pi \circ h(w)) + d_{M/\mathcal{D}}(\pi \circ h(w), \pi(M \smallsetminus W)) \\
&\leq \operatorname{diam}_{M/\mathcal{D}} \overline{U}_2 + d_{M/\mathcal{D}}(\pi \circ h(w), \pi(w)) \\
&< 2\delta + \varepsilon \\
&\leq 3\delta
\end{aligned}$$

which is a contradiction. Above we used that $\pi(\Delta), \pi \circ h(w) \in \overline{U}_2$, $\pi(w) \in \pi(M \smallsetminus W)$, and that $\varepsilon \leq \delta$. This contradiction implies that $h(w) \in M \smallsetminus \pi^{-1}(\overline{U}_2)$.

We continue with the outer *reductio ad absurdum*, namely showing that assuming $\overline{B} \not\subseteq W$ leads to a contradiction. From $h(w) \in M \smallsetminus \pi^{-1}(\overline{U}_2)$, it follows that $h(w) \notin \pi^{-1}(U_2) \smallsetminus \Delta$ and $h(w) \notin \pi^{-1}(U_1)$, since $(M \smallsetminus \pi^{-1}(\overline{U}_2)) \cap (\pi^{-1}(U_2) \smallsetminus \Delta) = \emptyset = (M \smallsetminus \pi^{-1}(\overline{U}_2)) \cap \pi^{-1}(U_1)$.

Next we consider the set B'. By the definition of $\mathcal{V}$, the set $\overline{B'}$ is contained in some element of $\mathcal{U}$. We also know that $h(w) \in \overline{B'}$. Then $\overline{B'} \not\subseteq \pi^{-1}(U_2) \smallsetminus \Delta$ since $h(w) \notin \pi^{-1}(U_2) \smallsetminus \Delta$. Similarly, $\overline{B'} \not\subseteq \pi^{-1}(U_1)$ since $h(w) \notin \pi^{-1}(U_1)$. Thus

$$\overline{B'} \subseteq M \smallsetminus \pi^{-1}(\overline{U}_2) \cup \pi^{-1}(U_3 \smallsetminus \overline{U}_1) = M \smallsetminus \pi^{-1}(U_1).$$

Since $h(\Delta) \subseteq B'$, for every $e \in \Delta$ we have that $h(e) \in B'$. We claim that additionally $h(e) \in \pi^{-1}(U_1)$. This follows since $d_{M/\mathcal{D}}(\pi(e), \pi \circ h(e)) < \varepsilon \leq \delta$ so that $\pi \circ h(e) \in U_1$ since $\pi(e) = \pi(\Delta)$.

So we have that

$$h(e) \in \pi^{-1}(U_1) \cap B' \subseteq \pi^{-1}(U_1) \cap M \smallsetminus \pi^{-1}(\overline{U}_1) = \emptyset.$$

Since this is impossible, we deduce that $\overline{B} \subseteq W$, so $\Delta \subseteq B \subseteq \overline{B} \subseteq W$ as desired. This completes the proof of Proposition 4.21. $\square$

CHAPTER 5

The Alexander gored ball and the Bing decomposition

Stefan Behrens and Min Hoon Kim

We introduce the *Alexander gored ball*, which is what we call the closure of the complement of the Alexander horned ball in S^3. We give three equivalent descriptions of the Alexander gored ball. In the final description, we present it as a decomposition space of D^3. By taking the double of the Alexander gored ball we obtain a decomposition space of S^3, with respect to the Bing decomposition that we saw in Example 4.6. We present Bing's proof that this decomposition shrinks, which implies that the double of the Alexander gored ball is homeomorphic to S^3. This instructive case study will introduce us to the subtleties of decompositions, particularly those with infinitely many non-singleton decomposition elements. Similar ideas appear here as in the upcoming 4-dimensional arguments, but being in three dimensions, visualisation is significantly easier. Moreover, the Bing decomposition is closely related to the gropes that will appear in Part II, and a variant of it will appear in the proof of the disc embedding theorem.

The idea of shrinking first arose in a paper of Bing [**Bin52**]. The main focus of Bing's paper [**Bin52**] was the Alexander horned sphere. In the 1930s Wilder [**Wil49**, Problem 4.6] had considered the question of whether the double of the exterior of the Alexander horned ball is homeomorphic to S^3. His interest arose from the fact that this doubled object has an obvious involution interchanging the two halves; if it turned out to be homeomorphic to S^3, this would give an interesting involution on the 3-sphere, namely one whose fixed point set is a very wild 2-sphere. Consequently, the topological involution would not be conjugate to a smooth involution. While he provided some evidence that his space was the 3-sphere, Wilder was unable to produce a conclusive proof and his question remained unanswered until Bing's paper.

5.1. Three descriptions of the Alexander gored ball

The Alexander horned sphere is shown in Figure 5.1. As depicted, the construction consists of iterating the following procedure. Starting at $n = 1$, in the nth step, first remove 2^{n+1} open discs from S^2. Attach 2^{n+1} annuli, along half of their boundary components, with linking between the middle pair of each set of four. Finally cap off the right hand boundaries of the 2^{n+1} annuli with 2^{n+1} smaller discs. Attach an infinite sequence of annuli and discs in this way, together with a final collection of points at the limit. This produces the Alexander horned sphere.

If instead we attach solid tubes to D^3, followed by the final set of limit points, the resulting object is a 3-dimensional ball topologically embedded into $\mathbb{R}^3$ and therefore into S^3. We call this the *Alexander horned ball*. Its boundary is the Alexander horned sphere.

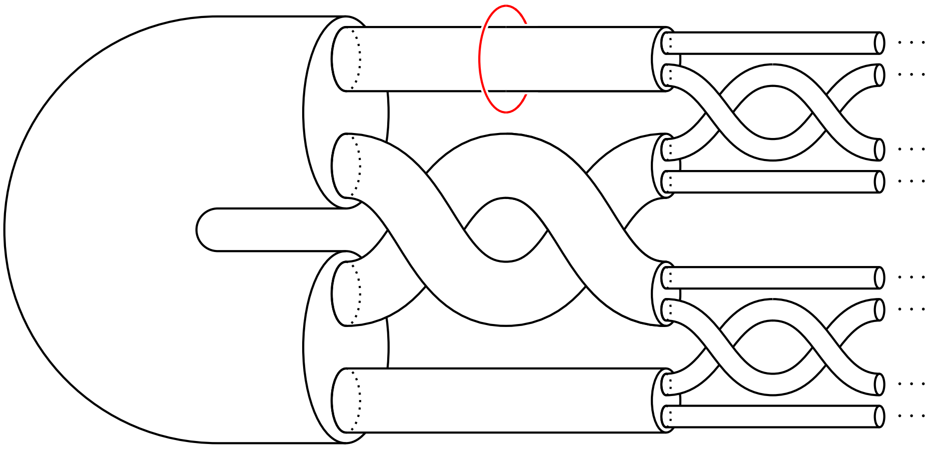


Figure 5.1. The Alexander horned sphere. A typical nontrivial element of the fundamental group of its complement is shown in red.

The closure of the complement of the Alexander horned ball in S^3 is what we call the *Alexander gored ball*, denoted by $\mathcal{A}$. We now give three descriptions of $\mathcal{A}$.

5.1.1. An intersection of 3-balls in D^3. The first description is an inside out version of Figure 5.1. Starting with the standard 3-ball, drill two pairs of holes, creating two "almost tunnels" which are "almost linked," as in the left picture of Figure 5.2. This space is still homeomorphic to a ball. The left two–thirds of Figure 5.1, with the eight skinniest tubes and everything to the right of them removed, should be compared with the left picture of Figure 5.2. We assert that the left picture of Figure 5.2 can be thought of as the left two–thirds of Figure 5.1 "turned inside–out." The two clasped tubes in the middle of Figure 5.1 correspond to the two clasped tubes in the left picture in Figure 5.2. The small stumps in the left picture of Figure 5.2 correspond to the top and bottom tubes of the middle four tubes in Figure 5.1.

Now repeat this construction infinitely many times, each time drilling twice as many almost tunnels as before, in the gaps where the previous tunnels were not quite completed. The next step is indicated on the right of Figure 5.2, which can be thought of as Figure 5.1 (minus the dots) turned inside–out. The space obtained after any finite number of iterations is still homeomorphic to a ball in the original D^3. The Alexander gored ball is the infinite intersection of these nested balls. However, note that this sequence of balls does not describe $\mathcal{A}$ as a cellular subset of D^3 since, in particular, none of the balls are contained in the interior of the previous ball in the sequence.

5.1.2. A (3-dimensional) grope. There is an equivalent picture of $\mathcal{A}$ as an infinite union of thickened, punctured tori with some limit points added in. The construction proceeds as follows. Let T be the 2-dimensional torus with an open disc removed and let $\mathbb{T} := T \times [0,1]$. We also fix a standard meridian-longitude pair of curves $\mu, \lambda \subseteq T$.

Start with a single copy $\mathbb{T}_0$ of $\mathbb{T}$ and attach two extra copies $\mathbb{T}_{00}$ and $\mathbb{T}_{01}$ along annular neighbourhoods of $\mu_0 \times \{0\}$ in $T \times \{0\}$ and of $\lambda_0 \times \{1\}$ in $T \times \{1\}$ respectively. Index the meridians and longitudes according to the corresponding copy of $\mathbb{T}$. Next,

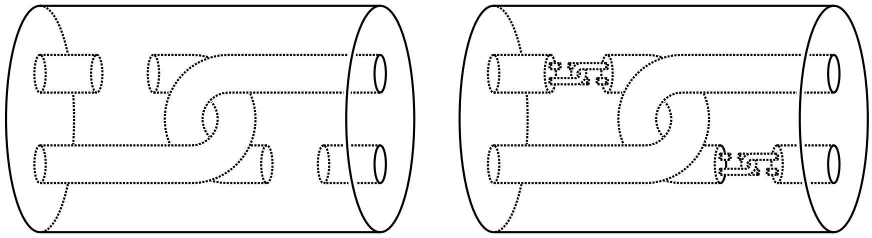


Figure 5.2. The Alexander gored ball as a countable intersection of 3-balls in D^3.

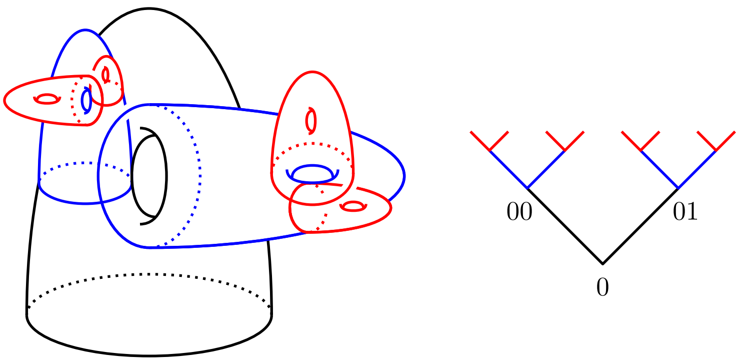


Figure 5.3. The Alexander gored ball as a grope. The 2-dimensional spine of the first three stages is shown on the left and the branching pattern is depicted as a tree on the right.

attach four more copies $\mathbb{T}_{000}$, $\mathbb{T}_{001}$, $\mathbb{T}_{010}$ and $\mathbb{T}_{011}$ along annular neighbourhoods of

$$\mu_{00} \times \{0\},\ \lambda_{00} \times \{1\},\ \mu_{01} \times \{0\} \text{ and } \lambda_{01} \times \{1\}$$

respectively. Iterate this procedure infinitely many times, each iteration adding twice as many copies of $\mathbb{T}$ as in the previous stage. The result is an infinite union of copies of $\mathbb{T}$ indexed by the vertices of a tree as indicated in Figure 5.3. The finite stages are examples of 3-dimensional *gropes*, and the infinite union is an infinite 3-dimensional grope.

This process can be done carefully so that the resulting space is embedded in 3-space, and in a way that forces the longitudes and meridians to get smaller and smaller in the successive stages so that they will ultimately converge to points. By inspecting the construction, these limit points form a Cantor set in 3-space. Indeed, they correspond to the limit points of the dyadic tree in Figure 5.3, which in turn correspond to infinite sequences of 0s and 1s.

We claim that this infinite union of thickened punctured tori together with the limit points is homeomorphic to $\mathcal{A}$. To see this we first look at D^3 carefully. Figure 5.4 shows a picture of $D^2 \times [-1, 1] \cong D^3$, and two thickened arcs embedded in $D^2 \times [-1, 1]$ in a clasp, much like in Figure 5.2, but here with the missing smaller plugs, also copies of $D^2 \times [-1, 1]$, temporarily put back in. The complement of these two thickened arcs is homeomorphic to $\mathbb{T}$. To assist with seeing this, in Figure 5.5 we show a copy of T in the complement of the thickened arcs, and it is not too hard to see that sweeping this left and right fills up all of the complement of the thickened arcs.

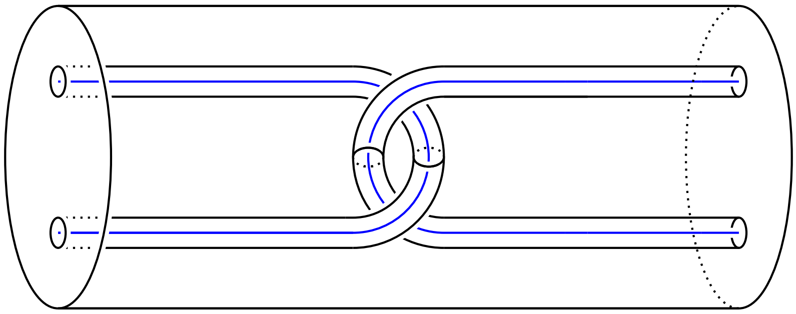


Figure 5.4. Two thickened arcs embedded in $D^2 \times [-1, 1]$.

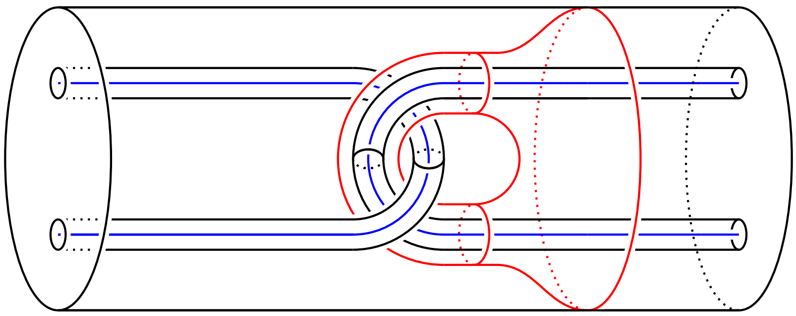


Figure 5.5. Two thickened arcs embedded in $D^2 \times [-1, 1]$. The complement of these thickened arcs is homeomorphic to $\mathbb{T}$. One such torus $T^2 \times \{0\} \subseteq \mathbb{T}$ is shown.

Now we take the picture of $D^2 \times [-1, 1]$ in Figure 5.4 and replace the thickened arcs, each of which is homeomorphic to $D^2 \times [-1, 1]$, with a copy of the entire picture, with no twisting. The resulting thickened arcs will look rather like Figure 5.6. Repeat this process infinitely many times.

Next, we describe the relationship between this and the previous construction of $\mathcal{A}$. On the one hand, we can modify this construction so that in each step we drill out thickened arcs (parametrised by $D^2 \times [-1, 1]$), but instead of gluing the model $D^2 \times [-1, 1]$ along the entire $S^1 \times [-1, 1]$, we glue a compressed smaller model $D^2 \times [-\varepsilon, \varepsilon]$ along $S^1 \times [-\varepsilon, \varepsilon]$, for some small $\varepsilon > 0$. This modified construction matches with Figure 5.2.

We call these plugs $D^2 \times [-\varepsilon, \varepsilon]$ *caps*, anticipating future language. Note that the caps become smaller and smaller with each step, and because their number doubles with each iteration, they converge to a Cantor set in D^3.

On the other hand, we know that removing the thickened arcs from $D^2 \times [-1, 1]$ in Figure 5.4 leaves us with a copy of $\mathbb{T} = T \times [0, 1]$. In each step of the construction, add in the plugs $D^2 \times [-\varepsilon, \varepsilon]$ and remove the new thickened arcs instead. This adds another stage of thickened punctured tori to the grope, and in the limit exhibits the infinite union of tori in Figure 5.3 embedded in D^3. In order to obtain the Alexander gored ball $\mathcal{A}$, we then have to add in the limit Cantor set of caps. So we have identified the construction of this subsection, of an infinite grope union limit points, with the description of $\mathcal{A}$ from Section 5.1.1 as an infinite intersection of nested 3-balls.

Remark 5.1. As mentioned, the union of tori construction given above is our first example of a *grope*. In this case, the grope is 3-dimensional. The grope is a thickened

2-complex consisting of a countable collection of surfaces glued together. We call the underlying 2-complex the *spine*. More general versions of this construction, where the thickenings of the spine are 4-dimensional, and where the torus can be replaced with an arbitrary orientable surface with a single boundary component, possibly a different such surface for every occurrence, will play a central rôle in the 4-dimensional arguments of Part II. See Chapter 12 for the precise definitions.

5.1.3. A decomposition space. Finally, we exhibit $\mathcal{A}$ as a decomposition space $D^3/\mathcal{D}$, where the thickened arcs in Figure 5.4 form the defining pattern of the decomposition $\mathcal{D}$. To construct the defining sequence, we glue in a copy of the original $D^2 \times [-1,1]$ in place of each thickened arc, and repeat this process infinitely many times. The thickened arcs in the second stage of this process are shown in Figure 5.6. The decomposition of $D^2 \times [-1,1] \cong D^3$ corresponding to this defining sequence is denoted by $\mathcal{D}$. Note that each cylinder in the defining sequence is contained in the topological interior of its antecedent, but not in the manifold interior.

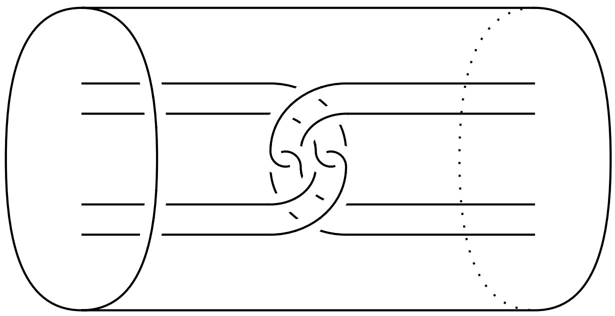


Figure 5.6. The second term in the defining sequence for a decomposition of D^3 giving rise to the Alexander gored ball. The defining pattern is given in Figure 5.4.

To see that the decomposition space $D^3/\mathcal{D}$ is homeomorphic to $\mathcal{A}$ we take a closer look at the first description of $\mathcal{A}$. There we had written $\mathcal{A} = \bigcap_{k=0}^{\infty} B_k$ as a countable intersection of 3-balls B_k such that $B_0 = D^3$ and $B_{k+1} \subseteq B_k$. The important observation is that B_{k+1} can be obtained from B_k by performing an ambient isotopy of 3-space. Indeed, this ambient isotopy may be taken to be fixed outside the small plugs mentioned in the previous section, namely the gaps between the "almost linked" tunnels. In particular, we have the homeomorphisms $h_k \colon B_k \to B_{k+1}$. Taking the limit

$$B_0 \xrightarrow{h_0} B_1 \xrightarrow{h_1} B_2 \xrightarrow{h_2} \cdots$$

of the compositions, we obtain a map $h_\infty \colon D^3 \to \mathcal{A}$ since $\bigcap_{k=0}^{\infty} B_k = \mathcal{A}$. It is not too hard to see that $h_\infty \colon D^3 \to \mathcal{A}$ is continuous.

The non-singleton inverse sets of h_∞ consist of all points that are moved by h_k for infinitely many k, and we assert that the maps h_k can be chosen so that these points agree with the elements of the decomposition $\mathcal{D}$.

Therefore $\mathcal{D}$ is given by the collection of nontrivial preimages of points of h_∞ and h_∞ induces a continuous bijection $D^3/\mathcal{D} \to \mathcal{A}$, which is a homeomorphism since $D^3/\mathcal{D}$ is compact and $\mathcal{A}$ is Hausdorff (as a subspace of the Hausdorff space D^3).

Remark 5.2. This construction of $\mathcal{A}$ brings up an interesting question about the Bing shrinking criterion. We have now described $\mathcal{A}$ as the result of crushing certain subsets in D^3 to points. This crushing could be done by a sequence of homeomorphisms of D^3 that do not move points too far in the quotient. So one might guess that $D^3/\mathcal{D}$ is homeomorphic to D^3 by the Bing shrinking criterion. However, such a sequence of maps does not satisfy the shrinking criterion.

Indeed, the fundamental group of $\operatorname{Int}\mathcal{A}$ is not trivial (see Remark 5.3) so the decomposition $\mathcal{D}$ is not shrinkable. A typical nontrivial element of the fundamental group is shown in Figure 5.1 in red.

One issue is that while the arcs to be crushed are an infinite intersection of nested cylinders, the cylinders are not *properly* nested, by which we mean that the connected components are not contained in the manifold interiors of the previous collection of cylinders, as mentioned above. Every cylinder in the defining sequence of cylinders touches the boundary ∂D^3 and these cylinders do not exhibit the corresponding decomposition of D^3 as a cellular decomposition.

However, when considering $D^3 \subseteq \mathbb{R}^3$, the shrinking of these subsets can be performed. This manifests itself in the fact that if we add a collar $S^2 \times [0,1]$ to $\mathcal{A}$, the objection of the previous paragraph disappears. In fact $\mathcal{A} \cup S^2 \times [0,1]$ is homeomorphic to D^3, as we see in the next remark. In particular, when we add a collar, we can augment the defining sequence of cylinders in the collar, to make them properly nested but without changing their infinite intersection, by adding cylinders of height $1/n$ in the nth step. Then apply Proposition 4.7 to make the non-singleton elements of the decomposition cellular. (Of course, not every cellular decomposition shrinks, so there is still something to show.) The 4-dimensional analogue of this will be crucial later and appears as the collar adding lemma (Lemma 25.1).

Remark 5.3. The fundamental group $\pi_1(\operatorname{Int}\mathcal{A})$ is nontrivial and perfect (see, for example, [**Hat02**, pages 170–171]). By applying [**Hat02**, Proposition 2B.1] to the embedding $h\colon D^3 \to S^3$ of the Alexander horned ball into S^3, one can see that $\widetilde{H}_k(\operatorname{Int}\mathcal{A}) = 0$ for all k. In particular, this shows that $\pi_1(\operatorname{Int}\mathcal{A})$ is perfect. On the other hand, the following collar argument (compare [**Bin64**, Theorem 4]) shows that $\mathcal{A}$ is contractible. Since $\mathcal{A} \cup_{\partial D^3} h(D^3 \smallsetminus \frac{1}{2}D^3)$ is the mapping cylinder of $h|_{\partial D^3}\colon \partial D^3 \to \mathcal{A}$, the space $\mathcal{A}$ is homotopy equivalent to $\mathcal{A} \cup_{\partial D^3} h(D^3 \smallsetminus \frac{1}{2}D^3)$. By the Schoenflies theorem,

$$\mathcal{A} \cup_{\partial D^3} h(D^3 \smallsetminus \frac{1}{2}D^3) \subseteq S^3$$

is homeomorphic to a 3-ball since it is bounded by the bicollared 2-sphere $h(\frac{1}{2}S^2)$. Therefore $\mathcal{A}$ is contractible. Note that $\mathcal{A}$ is *not* a topological manifold with boundary, so there is no contradiction in the fact that the topology of $\mathcal{A}$ and its interior differ so markedly.

5.2. The Bing decomposition: the first ever shrink

In this section we see the first decomposition that was ever shrunk, and how this was accomplished. Consider the quotient map $\pi\colon D^3 \to \mathcal{A}$ as in the previous section. We double the corresponding decomposition $\mathcal{D}$ of D^3 to obtain a decomposition $\mathcal{B} = D(\mathcal{D})$ of S^3 known as the *Bing decomposition* (see Example 4.6). The defining pattern for the Bing decomposition is given in Figure 5.7. Note that the corresponding decomposition space of S^3 is the double $\mathcal{A} \cup_{S^2} \mathcal{A} = D(\mathcal{A})$ of the Alexander gored ball. Here is Bing's remarkable theorem.

Theorem 5.4 ([**Bin52**]). *The Bing decomposition is shrinkable, the quotient map $\pi\colon S^3 \to D(\mathcal{A})$ is approximable by homeomorphisms, and thus $D(\mathcal{A})$ is homeomorphic to S^3.*

As mentioned earlier, this answered the question of Wilder whether the obvious involution on $D(\mathcal{A})$ is an exotic involution on S^3 instead of just an involution on some pathological metric space.

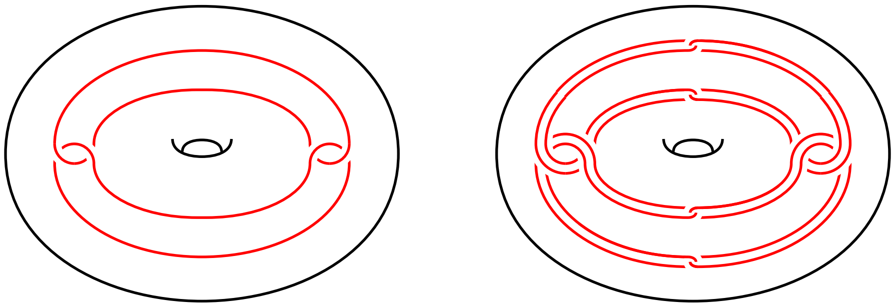


Figure 5.7. The defining pattern of the Bing decomposition (left) and its second stage (right).

Proof. By definition, the decomposition elements consist of the components of the countable intersection of nested solid tori where each solid torus contains two smaller solid tori forming the Bing double of the core. The nth stage of this construction yields 2^n solid tori, and in each stage the solid tori get thinner and thinner while the length of the core curves remains roughly the same. Let C be the core curve $S^1 \times \{0\}$ of the first and largest solid torus.

As a preliminary step, we must see that the Bing decomposition is upper semi-continuous. This is fairly immediate by Proposition 4.10. The sequence of solid tori in the defining sequence provide small saturated neighbourhoods for the decomposition elements, while any point not in any decomposition element is located in the complement of some smaller solid torus within a larger one, and we can find small saturated neighbourhoods therein.

Since the Bing decomposition is upper semi-continuous, the decomposition space $S^3/D(\mathcal{D})$ is metrisable and we fix some metric for the rest of the proof. In order to show that $D(\mathcal{D})$ is shrinkable, we fix $\varepsilon > 0$ and construct a self-homeomorphism of S^3 that shrinks each decomposition element to size less than ε with respect to the metric on S^3, and which does not move the decomposition elements too far away with respect to the chosen metric on $S^3/D(\mathcal{D})$, as in Definition 4.16.

The idea is as follows. We focus on a stage far enough along in the construction, stage n_ε of the defining sequence say, where the 2^{n_ε} solid tori are thinner than $\varepsilon/4$. More precisely, possibly after a sequence of ambient isotopies to carefully position the defining sequence (which by Proposition 4.7 does not change the homeomorphism type of the decomposition space), we assume that for sufficiently large n_ε each solid torus of stage n_ε lies in an $(\varepsilon/4)$-neighbourhood of the core circle C. In each of these solid tori we will produce an ambient isotopy that shrinks the decomposition elements.

For the purpose of producing shrinking homeomorphisms, we will measure the size of a decomposition element by bounding it above by the size of the corresponding solid torus in the n_ε stage of the defining sequence. We will measure size in terms of length along the core curve C of the first solid torus. By working on a small enough scale, sufficiently deep in the defining sequence, the curvature of C is assumed to be negligible. Therefore if we arrange for the image of each decomposition element under a shrinking homeomorphism (the outcome of the ambient isotopy, fixing the complement of the 2^{n_ε} solid tori) to have length at most $\varepsilon/2$, then by the triangle inequality the diameter of each element will be less than $\varepsilon/2 + \varepsilon/4 + \varepsilon/4 = \varepsilon$.

We assert that after possibly increasing the value of n_ε, any homeomorphism supported in the 2^{n_ε} solid tori will satisfy condition (i) of Definition 4.16 automatically. This will be proven more rigorously in Lemma 8.8, for now we give a brief sketch. It suffices to show that there is an n_ε such that each T' of the 2^{n_ε} solid tori has $\operatorname{diam} \pi(T') < \varepsilon$, since we only move points within a T', and thus every point gets moved less than ε as measured in the metric of $S^3/D(\mathcal{D})$, as required by (i). So, suppose for a contradiction that there is no such n_ε. Then there is a sequence of nested T' with the diameter of $\pi(T')$ greater than ε for every T'. But these T' limit to a decomposition element, so their image under π must have small diameter eventually, which leads to a contradiction. For the complete argument, see the proof of Lemma 8.8.

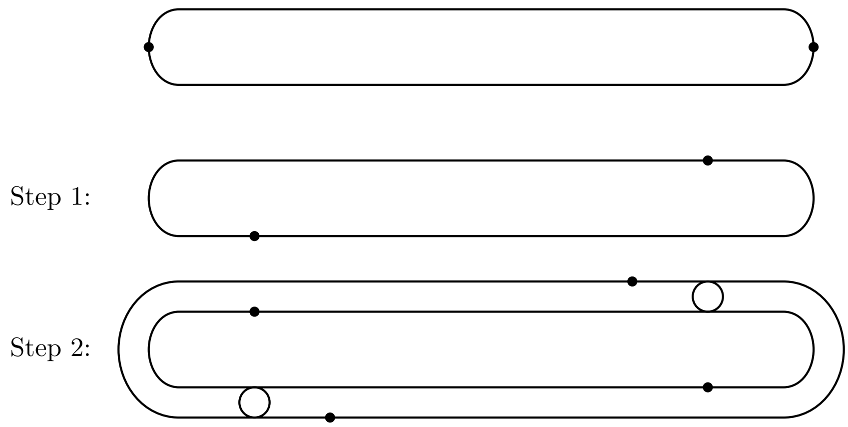


Figure 5.8. Bing's proof showing that the Bing decomposition shrinks. The dots in the middle picture are distance $\frac{1}{n}$ from the ends. Two of the dots in the bottom picture are distance $\frac{1}{n}$ from the end, and two of the dots are distance $\frac{1}{n}$ from the clasps.

It remains to address condition (ii) of Definition 4.16. Since all solid tori within a given stage are the same size and have the later solid tori embedded within them in the same way, it is enough to describe an ambient isotopy on a single solid torus S in stage n_ε that shrinks the decomposition elements within S. When applied to each of the 2^{n_ε} solid tori in stage n_ε, this gives rise to the desired homeomorphism, shrinking each $\Delta_i \in D(\mathcal{D})$ to diameter less than ε. Such a homeomorphism then will also satisfy condition (ii) of Definition 4.16.

Bing considered the segment of the core curve C near our solid torus S, that is the segment of C consisting of points of distance less than $\varepsilon/4$ from S, as a subset of the real line, and used distance along the real line to measure length. By performing ambient isotopies on the solid tori of the subsequent stages, we will make them have smaller and smaller length. As explained above, if a decomposition element has length less than $\varepsilon/2$ measured along the real line, then it has diameter less than ε.

Making solid tori have smaller length can be achieved by successively rotating the clasps as indicated in Figure 5.8, and described in more detail next. Suppose that the total length of the solid torus S shown in Figure 5.8, measured horizontally, is normalised to be 1. Let $\varepsilon' := \varepsilon/2E$, where E is the normalisation constant, that is the length of S using the metric on C induced from S^3. Let $n > n_\varepsilon$ be such that $\frac{1}{n} < \varepsilon'$. The clasps in the top picture of Figure 5.8 are represented by dots. Rotate the picture until the dots appear as in the second picture, so that the dots are distance $\frac{1}{n}$ from the corresponding ends. Now each component has length at most $1 - \frac{1}{n}$. This can be checked directly from the picture. The next rotation moves the clasps of the next stage, again by distance $\frac{1}{n}$. Now every component of the $(n_\varepsilon+1)$st stage has length at most $1-\frac{2}{n}$, which again can be checked directly. Repeat $n-1$ times in total on successively deeper stages, until every component in the $(n_\varepsilon+n-1)$th stage has length at most $1-\frac{n-1}{n}=\frac{1}{n}<\varepsilon'$. Let $\{T_i\}$ be the collection of solid tori that were moved in the final step i.e. the $(n_\varepsilon+n-1)$th stage tori. In the metric of S^3, prior to normalisation, each solid torus T_i has length less than $E\varepsilon'=\varepsilon/2$, and therefore has diameter less than ε.

The movement of solid tori described can be achieved by a self-homeomorphism of S^3 supported in the interior of the solid torus S. Apply the analogous self-homeomorphism to each of the 2^{n_ε} solid tori in stage n_ε, and extend by the identity outside the solid tori, to obtain a homeomorphism $h_\varepsilon\colon S^3\to S^3$. Let $\{T_i\}$ now denote the collection of all the $(n_\varepsilon+n-1)$ stage solid tori T_i, ranging over the stage n_ε solid tori labelled S above. Since every decomposition element is contained in some element of $\{T_i\}$, and $\operatorname{diam} T_i<\varepsilon$ for all i, it follows that $\operatorname{diam} h_\varepsilon(\Delta_i)<\varepsilon$ for every $\Delta_i\in D(\mathcal{D})$. Therefore condition (ii) of Definition 4.16 is satisfied. Recall that we already arranged for (i) of Definition 4.16 to hold by choosing n_ε sufficiently large. Therefore by Theorem 4.18 the Bing decomposition $D(\mathcal{D})$ shrinks, the quotient map $S^3\to S^3/D(\mathcal{D})$ is approximable by homeomorphisms, and in particular $S^3\cong S^3/D(\mathcal{D})$. This completes the proof of Theorem 5.4. $\square$

This shrinking procedure will be generalised in Chapter 8 to decompositions defined using arbitrary sequences of links in a solid torus, called *toroidal decompositions*. An alternative approach to shrinking the Bing decomposition was given in [**Bin88**].

CHAPTER 6

A decomposition that does not shrink

Stefan Behrens, Christopher W. Davis, and Mark Powell

So far we have seen both simple and complicated decompositions that shrink and soon we will expend much effort to show that certain other decompositions also shrink, often using only general properties of the decompositions rather than explicit constructions. To help the reader approach these proofs with a healthy dose of scepticism, we demonstrate now that not every decomposition shrinks. The content of this chapter is morally due to Bing [**Bin61**]. We give a slight modification of his construction and a more modern proof of non-shrinking.

The exposition in this chapter assumes some familiarity with 3-manifold topology and with the theory of knots and links in S^3 in particular. A general analysis of when toroidal decompositions of the sort considered here and in the previous chapter shrink can be found in [**KP14**], a summary of which appears in Chapter 8.

We consider the decomposition $\mathcal{B}_2$ of S^3 determined by the defining pattern given in Figure 6.1. Similarly to the original Bing decomposition, the decomposition $\mathcal{B}_2$ consists of the connected components of an infinite intersection of solid tori which are nested in such a way that, at each stage, two of them are placed inside each solid torus from the previous stage. The only difference is in how the two are placed. The goal of this chapter is to prove the following.

THEOREM 6.1 ([**Bin61**]). *The decomposition $\mathcal{B}_2$ does not shrink.*

We call the two nested solid tori in the defining pattern P and Q and consider two generic meridional discs A and B in the very first stage solid torus T. Fixed a choice of metric on the quotient $S^3/\mathcal{B}_2$. We will show that for any homeomorphism $h\colon S^3 \to S^3$ satisfying the first condition of Definition 4.16, namely that points are not moved very far the metric of $S^3/\mathcal{B}_2$, the image of some decomposition element under h has to intersect both A and B. By choosing A and B to be far apart, this will guarantee the existence of some large decomposition element, which will in turn contravene the second condition of Definition 4.16.

More precisely, we will shift perspective slightly, and work with discs

$$A_0 := h^{-1}(A) \text{ and } B_0 := h^{-1}(B)$$

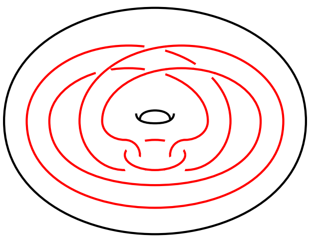


Figure 6.1. The defining pattern for the decomposition $\mathcal{B}_2$.

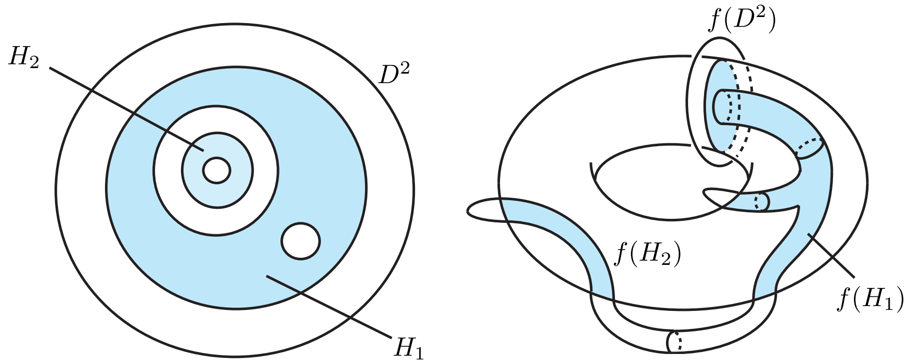


Figure 6.2. On the right we see the image of a map $f\colon D^2 \to S^3$ and a solid torus $M \subseteq S^3$. The inverse image $f^{-1}(M) \subseteq D^2$ has two connected components, H_1 and H_2, shown on the left. The restriction $f|_{H_1}$ is substantial while the restriction $f|_{H_2}$ is not. Since $f|_{H_1}$ is substantial, by definition $f(D^2)$ has a substantial intersection with M.

that depend on the homeomorphism h, and we will investigate how these discs intersect the nested solid tori in the defining sequence for $\mathcal{B}_2$.

To show that for every $h\colon S^3 \to S^3$ as above there is always a decomposition element intersecting both A_0 and B_0, we will employ the language of *substantial intersections*. This gives a notion of intersections of the discs A_0 and B_0 with the solid tori in the defining sequence that can be used in an induction; requiring meridional disc intersections instead would be too rigid.

Definition 6.2 (Substantial intersections).

(1) Let H be a compact, connected 2-dimensional submanifold of a disc D^2, i.e. a planar surface. Let D_H denote the disc in D^2 such that ∂D_H is equal to the outermost boundary component of H. Let M be a compact 3-manifold with boundary. A map $g\colon H \to M$ is called *substantial* if $g(\partial H) \subseteq \partial M$ and g extends to a map $G\colon D_H \to M$ with $G(D_H \smallsetminus H) \subseteq \partial M$ but $g|_{\partial D_H}$ does not extend to a map of D_H into ∂M.

(2) Let $f\colon D^2 \to S^3$ be a proper map of a disc D^2 into S^3. Let M be a submanifold of S^3 homeomorphic to the solid torus $S^1 \times D^2$. We say that $f(D^2)$ has a *substantial intersection* with M if on some connected component H of $f^{-1}(M)$, the map $g := f|_H \colon H \to M$ is substantial.

Figure 6.2 shows an example of a substantial intersection. For the next lemma, which gives a useful characterisation of substantial intersections, recall that a meridian of $M \cong S^1 \times D^2$ is a curve on ∂M isotopic to $\{p\} \times \partial D^2$ for some $p \in S^1$.

Lemma 6.3. *Let $M \subseteq S^3$ be a solid torus. Let $f\colon D^2 \to S^3$ be a map of the disc such that $f^{-1}(\partial M)$ consists of mutually disjoint simple closed curves and such that the restriction $f|_{f^{-1}(\partial M)}$ is an embedding. Then $f(D^2)$ has substantial intersection with M if and only if the image of f contains a meridian for M.*

Proof. First assume that there is a curve $\delta \subseteq D^2$ such that $f(\delta)$ is a meridian for M. Now, any other simple closed curve in ∂M which is also in the image of f must be disjoint from $f(\delta)$. Since the result of cutting ∂M open along δ is an annulus, the only such isotopy classes are parallels of $f(\delta)$ or curves that bound discs in ∂M.

Let $\{\delta_i\}_{i=1}^n$ be the set of all curves in D^2 sent by f to meridians for M. Each of these bound discs in D^2. Let D' be an innermost such disc and let δ' be its boundary curve. It follows that δ' must be an outermost boundary component of some component H of $f^{-1}(M)$. We claim that the restriction $f|_H$ is substantial. First notice that by the assumption that D' be an innermost disc bounded by a curve sent to a meridian for M, every other boundary component of H must bound a disc in ∂M. Thus $f|_H$ extends to a map from D' to M sending $D' \smallsetminus H$ to ∂M. Secondly, since the meridian is essential in $\pi_1(\partial M)$, there is no extension of $f|_{\partial H}$ to a map $D' \to \partial M$. This proves that $f(D^2)$ has substantial intersection with M.

Conversely, suppose that there is no curve in D^2 sent by f to a meridian for M. In the case that there is no curve δ in D^2 sent by f to an essential curve in ∂M, every boundary component of every $H \subseteq f^{-1}(M)$ bounds a disc in ∂M. From here it follows from the definition that f has no substantial intersection with M.

Let γ then be a curve in D^2 sent by f to an essential curve $f(\gamma)$ in ∂M. As before we then conclude that every curve in $f^{-1}(\partial M)$ is sent either to a parallel of $f(\gamma)$ or to a curve that bounds a disc in ∂M. If $f(D^2)$ has substantial intersection with M, then let H be a component of $f^{-1}(M)$ such that $f|_H$ is substantial in M. It would follow that the outermost boundary component of H does not bound a disc in ∂M, and so must be a parallel of $f(\gamma)$. Moreover every other boundary component bounds a disc in ∂M. It follows that $f(\gamma)$ bounds a disc in M. But the only essential curve in the boundary of a solid torus which bounds a disc interior to that solid torus is a meridian. This contradiction completes the proof. $\square$

The following proposition shows that given any two discs bounded by meridians for T, one of the solid tori P and Q will have substantial intersection with both discs. As explained above, we will use this to show that some decomposition element has to be large, which will lead to the impossibility of finding a homeomorphism that satisfies both conditions of Definition 4.16.

Proposition 6.4. *Let $P, Q \subseteq T$ denote the solid tori in Figure 6.1, where T is the standard unknotted solid torus in S^3. Let $A, B \colon D^2 \to S^3$ be two discs in S^3 bounded by meridians for T. Then one of P or Q has substantial intersection with both $A(D^2)$ and $B(D^2)$.*

Proof. We will often abuse notation by using A and B to refer to the images $A(D^2)$ and $B(D^2)$ respectively.

Pass to the standard (nontrivial) 2-fold cover $\widetilde{T}$ of the solid torus T. The solid tori P and Q have two disjoint lifts each. Let $\widetilde{P}$ and $\widetilde{Q}$ respectively denote lifts of P and Q as shown in Figure 6.3. Observe that in the lifted picture the solid tori $\widetilde{P}$ and $\widetilde{Q}$ are in the same position as in the case of the Bing decomposition that we studied in the previous section.

As in the proof of Lemma 6.3, the intersection between A and ∂T consists of parallel copies of the meridian of T and simple closed curves that bound discs in ∂T. Within the domain D^2 of A, consider the simple closed curves mapping to meridians of T. Choose an innermost such curve, namely one that bounds a disc Δ in D^2 containing no other curve mapped to a meridian of T under A. Then $A(\Delta)$ has substantial intersection with T by Lemma 6.3. In particular, modifying the map A, by Definition 6.2 there is an extension $A \colon \Delta \to T$, such that the subsets of D^2 on which A was modified now map to ∂T. Therefore a substantial intersection of $A(\Delta)$ with P or Q implies a substantial intersection of the original A with P or Q respectively.

We also see that $A(\Delta)$ has two lifts to $\widetilde{T}$, call them $\widetilde{A}_1$ and $\widetilde{A}_2$. Similarly, define $\widetilde{B}_1$ and $\widetilde{B}_2$. Observe that $\widetilde{A}_1$, $\widetilde{A}_2$, $\widetilde{B}_1$, and $\widetilde{B}_2$ are all meridional discs of $\widetilde{T}$. Let $\widetilde{A} := \widetilde{A}_1 \cup \widetilde{A}_2$ and $\widetilde{B} := \widetilde{B}_1 \cup \widetilde{B}_2$.

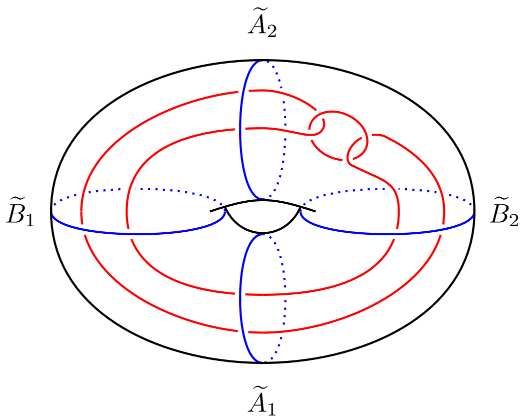


Figure 6.3. Lifts $\widetilde{P}$ and $\widetilde{Q}$ (red) of the components of the defining pattern together with the curves bounding the lifts $\widetilde{A}_1$, $\widetilde{B}_1$, $\widetilde{A}_2$, and $\widetilde{B}_2$ (blue).

By transversality, we can perform an isotopy of the boundaries of P and Q so that the intersection of $\partial\widetilde{P}$ and $\partial\widetilde{Q}$ with $\widetilde{A}$ and $\widetilde{B}$ consists of a finite union of disjoint simple closed curves. Such an isotopy, if suitably small and supported near ∂P and ∂Q, does not change the decomposition elements of $\mathcal{B}_2$.

LEMMA 6.5. *In $\widetilde{T}$, $\widetilde{A}_1$ has substantial intersection with one of the solid tori $\widetilde{P}$ or $\widetilde{Q}$. By symmetry, the same is also true for $\widetilde{A}_2$, $\widetilde{B}_1$, and $\widetilde{B}_2$.*

PROOF. Let $\widetilde{p} \subseteq \widetilde{P}$ and $\widetilde{q} \subseteq \widetilde{Q}$ be the cores $S^1 \times \{0\}$ of the solid tori $S^1 \times D^2$. Suppose that the intersection between $\widetilde{A}_1$ and $\widetilde{P}$ is not substantial. By Lemma 6.3 it follows that $\partial\widetilde{P} \cap \widetilde{A}_1$ contains no meridians for $\widetilde{P}$. Now let $\gamma_1, \dots, \gamma_k$ be the components of $\partial\widetilde{P} \cap \widetilde{A}_1$. If some γ_j bounds a disc in $\partial\widetilde{P}$ disjoint from all other γ_j then we modify $\widetilde{A}_1$ using this disc, reducing the number of intersections with $\partial\widetilde{P}$. Iterating this process, we obtain a disc A_1', and thus reduce to the setting that $\partial\widetilde{P} \cap A_1'$ consists of a collection of parallel copies of some essential simple closed curve γ on $\partial\widetilde{P}$.

One of these curves bounds an innermost disc in A_1' disjoint from the other curves. As a consequence, since all the curves are parallel copies of γ, we see that γ bounds a disc in the complement of $\partial\widetilde{P}$. The only two such curves are the meridian and longitude. Since there is no substantial intersection by assumption, the meridian is precluded. Thus A_1' contains a longitude of $\widetilde{P}$. After isotoping the core $\widetilde{p}$ of $\widetilde{P}$ we arrange for $\widetilde{p}$ to be this longitude and so lie on A_1'. But then by a small push we can displace $\widetilde{p}$ off A_1', to arrange for $\widetilde{p}$ and A_1' to be disjoint.

Apply the same procedure above now to A_1' and $\widetilde{Q}$. We still call the final disc A_1'.

Thus if both $\widetilde{P}$ and $\widetilde{Q}$ have no substantial intersection with $\widetilde{A}_1$, then some meridian of $\widetilde{T}$, namely the boundary of A_1', would bound a disc disjoint from curves isotopic to the cores $\widetilde{p}$ and $\widetilde{q}$. However, the cores $\widetilde{p}$ and $\widetilde{q}$ of $\widetilde{P}$ and $\widetilde{Q}$ along with a meridian of $\widetilde{T}$ form the Borromean rings. We have then asserted that one component of the Borromean rings bounds a disc in the complement of the remaining two components. Since any two components of the Borromean rings form the unlink, this would then imply that the Borromean rings are unlinked, which is false, as seen, for example, from the nontriviality of the Milnor triple linking number

$\overline{\mu}_L(123)$ of the Borromean rings L. Here the Milnor triple linking number is an integer valued invariant of 3-component linking number zero links that vanishes on the unlink [**Mil57**].

This contradiction completes the proof that the disc $\widetilde{A}_1$ has a substantial intersection with one of $\widetilde{P}$ and $\widetilde{Q}$. As in the statement of the lemma, by symmetry each of $\widetilde{A}_2$, $\widetilde{B}_1$, and $\widetilde{B}_2$ has the same property. □

Now we continue with the proof of Proposition 6.4. We prove that one of P and Q has substantial intersection with both A and B.

Observe that if A does not have substantial intersection with P, then in the cover each of $\widetilde{A}_1$ and $\widetilde{A}_2$ does not have substantial intersection with $\widetilde{P}$. According to Lemma 6.5, they must then both have substantial intersection with $\widetilde{Q}$.

Now if $\widetilde{Q}$ has no substantial intersection with either of $\widetilde{B}_1$ and $\widetilde{B}_2$, then as in the proof of Lemma 6.5, $\widetilde{Q}$ could be taken disjoint from a pair of discs D, D' with the same boundary as $\widetilde{B}_1$ and $\widetilde{B}_2$ and such that D and D' are both disjoint from $\widetilde{A}_1$ and $\widetilde{A}_2$. But these discs D and D' would then separate the solid torus $\widetilde{T}$ into two components such that $\widetilde{A}_1$ and $\widetilde{A}_2$ lie in different components. Since $\widetilde{Q}$ has substantial intersection with both $\widetilde{A}_1$ and $\widetilde{A}_2$, the core $\widetilde{q}$ of $\widetilde{Q}$ gives a path from a point in $\widetilde{A}_1$ to a point in $\widetilde{A}_2$, which must intersect one of D or D'. This contradicts that $\widetilde{Q}$ is disjoint from D and D'. Therefore $\widetilde{Q}$ has substantial intersection with one of $\widetilde{B}_1$ or $\widetilde{B}_2$.

We see that if A does not have substantial intersection with P then both of $\widetilde{A}_1$, $\widetilde{A}_2$, and at least one of $\widetilde{B}_1$ and $\widetilde{B}_2$, have substantial intersection with $\widetilde{Q}$. Pushing these intersections downstairs we conclude that A and B both have substantial intersection with Q, completing the proof of Proposition 6.4. □

Now we are ready to show that the decomposition $\mathcal{B}_2$ does not shrink.

Proof of Theorem 6.1. Recall that we write T for the initial solid torus in the defining sequence for $\mathcal{B}_2$, and we call the two nested solid tori in the defining pattern P and Q, shown in Figure 6.1. Let A and B be embedded meridional discs in T, union thin annuli of width $\delta > 0$ in $S^3 \smallsetminus \operatorname{Int} T$. Suppose that A and B are distance $\kappa > 0$ apart. Choose

$$\varepsilon < \min\{\kappa, \delta/2\}.$$

Write $\pi\colon S^3 \to S^3/\mathcal{B}_2$ for the quotient map. We want to choose a metric on $S^3/\mathcal{B}_2$, which we may do since $\mathcal{B}_2$ is upper semi-continuous (Corollary 4.13). Hausdorff [**Hau30**] showed that for a metrisable space E with a subspace F, and prescribed metric on F, one may find a metric on E that extends the given metric on F. Apply this to $S^3/\mathcal{B}_2$, to obtain a metric that agrees with the standard metric on $\overline{S^3 \smallsetminus T}$. In other words, for all $x, y \in \overline{S^3 \smallsetminus T}$, we have $d_{S^3}(x,y) = d_{S^3/\mathcal{B}_2}(\pi(x), \pi(y))$. We have used that $\pi|_{\overline{S^3 \smallsetminus T}}$ is a homeomorphism.

Now let $h\colon S^3 \to S^3$ be a homeomorphism such that $d_{S^3/\mathcal{B}_2}(\pi(x), \pi(h(x))) < \varepsilon$ for every $x \in S^3$, that is h satisfies (i) of Definition 4.16. Then by the construction of the metric $d_{S^3/\mathcal{B}_2}$, for every $x \in S^3$ whose distance from T is at least ε, both $x, h(x) \in \overline{S^3 \smallsetminus T}$ and so $d_{S^3}(h(x), x) < \varepsilon$.

For $\nu > 0$, let $N_\nu(T)$ denote a ν-neighbourhood of T, namely all the points of distance at most ν from T.

Claim. *$h^{-1}(\partial A) \subseteq N_{3\delta/2}(T) \smallsetminus T$ and $h^{-1}(\partial B) \subseteq N_{3\delta/2}(T) \smallsetminus T$. Moreover, both curves have linking number one with the core of T.*

The first sentence is immediate from the properties of h and the choice of ε and δ. The homeomorphism h moves points of ∂A less than ε. The linking number between ∂A (respectively $h^{-1}(\partial A)$) and the core of T can be computed as the degree of the Gauss map $S^1 \times S^1 \to S^2$ given by sending (x, y), with $x \in \partial A$ (respectively $h^{-1}(\partial A)$) and y in the core of T, considered as vectors in $\mathbb{R}^3$, to $(x-y)/\|x-y\|$. But a perturbation of the Gauss map, small relative to the diameter of S^2, does not change its degree. This proves the claim.

We next prove that some decomposition element intersects both of the discs $A_0 := h^{-1}(A)$ and $B_0 := h^{-1}(B)$. By the claim, and linking number considerations, both A_0 and B_0 have substantial intersections with T. In more detail, ∂A_0 and the core of T have linking number one, so form a nontrivial link. The argument of Lemma 6.5, with the linking number used instead of the triple linking number, shows that A_0 and B_0 has a substantial intersection with T, and similarly for B. By Lemma 6.3, applied after making a small isotopy of ∂T so that A_0 and ∂T intersect transversely (we can make such an isotopy without affecting the decomposition elements), A_0 contains a meridian of T and therefore contains a smaller disc A_0' with boundary a meridian of T. Similarly B_0 contains a smaller disc B_0' with boundary a meridian of T. By Proposition 6.4, one of P or Q has substantial intersection with both A_0' and B_0', and therefore with A_0 and B_0. Without loss of generality let it be P.

By Lemma 6.3, A_0 and B_0 both contain discs A_1 and B_1 in S^3 bounded by meridians for P. Let $P_1, Q_1 \subseteq P$ be the solid tori of the next stage of the defining sequence for $\mathcal{B}_2$. Apply Proposition 6.4 again to see that one of P_1 or Q_1 has substantial intersection with both of A_1 and B_1, and so has substantial intersection with both A_0 and B_0. Again we assume without loss of generality that P_1 has these substantial intersections.

Iterate this procedure to obtain a sequence $P \supsetneq P_1 \supsetneq P_2 \supsetneq \cdots$ in the defining sequence with substantial (and in particular nonempty) intersection with A_0 and B_0. Thus the decomposition element $\Delta = P \cap P_1 \cap P_2 \cap \cdots$ intersects both $A_0 = h^{-1}(A)$ and $B_0 = h^{-1}(B)$. This implies that $h(\Delta)$ intersects both A and B. Therefore $\operatorname{diam}_{S^3} h(\Delta)$ is at least as large as the minimal distance from a point in A to a point in B, that is $\operatorname{diam} h(\Delta) > \kappa > \varepsilon$.

Now $h \colon S^3 \to S^3$ was chosen to be some homeomorphism satisfying (i) of Definition 4.16. We have shown that there exists an ε such that for any such h, (ii) of Definition 4.16, that $\operatorname{diam}_{S^3} h(\Delta) < \varepsilon$ for every decomposition element $\Delta \in \mathcal{B}_2$, is not satisfied. Therefore $\mathcal{B}_2$ does not shrink. This completes the proof of Theorem 6.1. $\square$

CHAPTER 7

The Whitehead decomposition

Xiaoyi Cui, Boldizsár Kalmár, Patrick Orson, and Nathan Sunukjian

We study the Whitehead decomposition of S^3. In particular, we will show that the Whitehead decomposition of S^3 does not shrink but that the product of the Whitehead decomposition with the trivial decomposition of the real line does shrink. In contrast to the other chapters in Part I, the results of this chapter will not be directly used in the proof of the disc embedding theorem in Part IV. This is due to our decision to use skyscrapers instead of Casson handles, which were the main tool in Freedman's original proof [**Fre82a**]. The beautiful ideas in this chapter have been included for anyone trying to understand Freedman's original proof and for the insights they give into shrinkability. Moreover, in the next chapter we will consider mixed ramified Bing-Whitehead decompositions and while the main theorem of this chapter is not used directly later on, it will provide some important context to the results that we do use.

The *Whitehead decomposition*, denoted by $\mathcal{W}$, is the decomposition of S^3 described by the defining sequence of nested solid tori with defining pattern as in Figure 7.1. In other words, we start with the standard unknotted solid torus $T_0 \subseteq S^3$ and in each stage $i \geq 1$, a solid torus T_i is embedded into its predecessor as a regular neighbourhood of the untwisted *Whitehead double* of the core. The clasp of the Whitehead double does not matter for us and can in general differ at each stage. However, in order to make a precise definition, we use the negative clasp depicted in Figure 7.1. The Whitehead decomposition $\mathcal{W}$ is then defined to be the collection of connected components of the infinite intersection $\bigcap_{i=0}^{\infty} T_i$. Since each T_i is nonempty, connected, and compact, and the T_i are nested, the next lemma implies that $\mathcal{W}$ is connected.

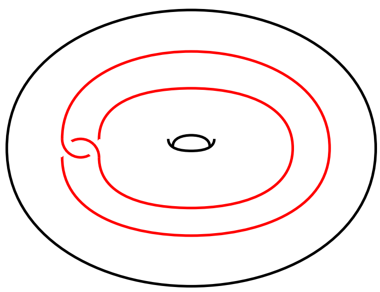


Figure 7.1. The defining pattern for the Whitehead decomposition.

Lemma 7.1. *Let $X_1, X_2, \dots$ be a sequence of nonempty, compact, connected subsets of a compact metric space X with $X_{i+1} \subseteq X_i$ for all i. Then $\bigcap_{i=1}^{\infty} X_i$ is nonempty and connected.*

PROOF. First we show that $\bigcap_{i=1}^{\infty} X_i$ is nonempty. Suppose for a contradiction that $\bigcap_{i=1}^{\infty} X_i$ is empty. As a compact subset of a metric space, each X_i is closed. Then $\{X \smallsetminus X_i\}_{i=1}^{\infty}$ is an open cover of X, so has a finite subcover, whose union is $X \smallsetminus X_j$ for some j. But $X_j \neq \emptyset$, so $X \smallsetminus X_j \neq X$, and we have a contradiction, completing the proof that $\bigcap_{i=1}^{\infty} X_i$ is nonempty.

Now, suppose that $\bigcap_{i=1}^{\infty} X_i$ is not connected. Let $A, B \subseteq \bigcap_{i=1}^{\infty} X_i$ be disjoint, closed, nonempty subsets with $A \cup B = \bigcap_{i=1}^{\infty} X_i$. Let $U \supseteq A$ and $V \supseteq B$ be open sets in X with $U \cap V = \emptyset$. These exist since X is a compact metric space, and A and B have positive distance between them, as disjoint compact subsets.

Let

$$F_i := X_i \smallsetminus (U \cup V)$$

for each i. Since U and V are open, this is a closed subset of X_i, which is compact, so each F_i is compact. Now the $\{F_i\}_{i=1}^{\infty}$ form a nested sequence of compact sets with

$$\bigcap_{i=1}^{\infty} F_i = \bigcap_{i=1}^{\infty} (X_i \smallsetminus (U \cup V)) = \big(\bigcap_{i=1}^{\infty} X_i\big) \smallsetminus (U \cup V) = \emptyset.$$

By Cantor's intersection theorem [**Mun00**, Theorem 26.9], $F_j = \emptyset$ for some j. It follows that $X_j \subseteq U \cup V$.

We claim that X_j intersects both U and V. Suppose for a contradiction and without loss of generality that $X_j \subseteq U$. Then $\bigcap_{i=1}^{\infty} X_i \subseteq X_j \subseteq A$. But also $\bigcap_{i=1}^{\infty} X_i = A \cup B$, so

$$B = (A \cup B) \smallsetminus A = (\bigcap_{i=1}^{\infty} X_i) \smallsetminus A = \emptyset$$

which contradicts that B is nonempty. Thus X_j intersects both U and V as claimed.

Define $U_j := X_j \cap U$ and $V_j := X_j \cap V$. Both of these are nonempty open subsets of X_j, and we have that $U_j \cup V_j = X_j \cap (U \cup V) = X_j$ since $X_j \subseteq U \cup V$, and $U_j \cap V_j \subseteq U \cap V = \emptyset$. This contradicts that X_j is connected, completing the proof that $\bigcap_{i=1}^{\infty} X_i$ is connected. □

7.1. The Whitehead decomposition does not shrink

Unlike the Bing decomposition $\mathcal{B}$ and the decomposition $\mathcal{B}_2$ of Chapters 5 and 6, each of which had infinitely many nontrivial elements, by Lemma 7.1 $\mathcal{W}$ has a single nontrivial element, which we also denote by the symbol $\mathcal{W}$ by an abuse of notation. We have previously studied decompositions with finitely many nontrivial elements in Chapter 3. As we saw in Proposition 4.21, if an upper semi-continuous decomposition of a manifold is shrinkable then each element of the decomposition must be cellular. It is clear from the construction that the Whitehead decomposition is upper semi-continuous by Proposition 4.10 since, like for the Bing decomposition, the sequence of solid tori in the defining sequence provide arbitrarily small saturated neighbourhoods for the decomposition element, while any point not in the decomposition element is located in the complement of a smaller solid torus within a larger one and we can find small saturated neighbourhoods therein. Similarly, it is straightforward to establish that the single element of $\mathcal{W}$ is not cellular and thus the Whitehead decomposition does not shrink. Alternatively, the reader might prefer to argue directly as we did in Chapter 6. We will soon prove the following theorem.

THEOREM 7.2. *The decomposition space $S^3/\mathcal{W}$ is not a manifold.*

Since the decomposition space is thus, in particular, not homeomorphic to S^3, this gives yet another means to see that that the Whitehead decomposition does not shrink. We will need the following definition.

DEFINITION 7.3. A noncompact space X is said to be *simply connected at infinity* if for all compact $K \subseteq X$ there exists a compact L with $K \subseteq L \subseteq X$ such that the inclusion induced map $\pi_1(X \setminus L) \to \pi_1(X \setminus K)$ is trivial.

PROPOSITION 7.4. *The space $S^3 \setminus \mathcal{W}$ is not simply connected at infinity.*

We will prove this proposition after explaining how it implies Theorem 7.2.

PROOF OF THEOREM 7.2. We show that if $S^3/\mathcal{W}$ were a manifold, then the space $X = S^3 \setminus \mathcal{W}$ would be simply connected at infinity.

Let $\pi\colon S^3 \to S^3/\mathcal{W}$ be the quotient map and suppose that $S^3/\mathcal{W}$ is a manifold. Then in particular $y := \pi(\mathcal{W})$ is a manifold point. Given any compact set $K \subseteq S^3 \setminus \mathcal{W}$, there exists a neighbourhood $B \subseteq S^3/\mathcal{W}$ of y which is homeomorphic to the open 3-ball and disjoint from $\pi(K)$. Here we are using the fact that $S^3/\mathcal{W}$ is Hausdorff, which follows from Proposition 4.12 since S^3 is Hausdorff and $\mathcal{W}$ is upper semi-continuous. Then let $L := (S^3 \setminus \mathcal{W}) \setminus \pi^{-1}(B \setminus \{y\})$. By construction, $K \subseteq L$. Next we show that L is compact. To see this, note that

$$L = (S^3 \setminus \mathcal{W}) \setminus \pi^{-1}(B \setminus \{y\}) = \pi^{-1}(S^3/\mathcal{W} \setminus B).$$

Then $S^3/\mathcal{W} \setminus B$ is a closed subset of the compact space $S^3/\mathcal{W}$, so is compact. Moreover, since $S^3/\mathcal{W} \setminus B \subseteq S^3/\mathcal{W}$ is closed, the map $\pi|_{\pi^{-1}(S^3/\mathcal{W} \setminus B)}$ is a closed map and since it is bijective it is a homeomorphism. This implies that $L = \pi^{-1}(S^3/\mathcal{W} \setminus B)$ is compact.

Finally, we need to show that the map $\pi_1((S^3 \setminus \mathcal{W}) \setminus L) \to \pi_1((S^3 \setminus \mathcal{W}) \setminus K)$ is trivial. It will suffice to show that the space $(S^3 \setminus \mathcal{W}) \setminus L$ is simply connected. The key point is that since $B \setminus \{y\}$ is an open subset of $S^3/\mathcal{W}$, the map $\pi|_{\pi^{-1}(B \setminus \{y\})}$ is a bijective open map and so is a homeomorphism. Then note that

$$\begin{aligned}(S^3 \setminus \mathcal{W}) \setminus L &= (S^3 \setminus \mathcal{W}) \setminus ((S^3 \setminus \mathcal{W}) \setminus \pi^{-1}(B \setminus \{y\})) \\ &= \pi^{-1}(B \setminus \{y\}) \\ &\cong B \setminus \{y\}.\end{aligned}$$

The space $B \setminus \{y\}$ is homotopy equivalent to S^2, which is simply connected. We have shown that if $S^3/\mathcal{W}$ were a manifold, then $S^3 \setminus \mathcal{W}$ would be simply connected at infinity. Proposition 7.4 then completes the proof. □

PROOF OF PROPOSITION 7.4. Let $K_i := S^3 \setminus \operatorname{Int} T_i$. It will suffice to show that for all compact $K' \supseteq K_0$, the inclusion induced map $\pi_1((S^3 \setminus \mathcal{W}) \setminus K') \to \pi_1((S^3 \setminus \mathcal{W}) \setminus K_0)$ is nontrivial. Suppose for a contradiction that this is false. That is, there exists some $K' \supseteq K_0$ such that the inclusion induced map $\pi_1((S^3 \setminus \mathcal{W}) \setminus K') \to \pi_1((S^3 \setminus \mathcal{W}) \setminus K_0)$ is trivial. Since $\{K_i\}$ is an exhaustion of $S^3 \setminus \mathcal{W}$ by compact sets, that is $S^3 \setminus \mathcal{W} = \bigcup_{i=0}^{\infty} K_i$, there is some $i \geq 1$ such that $K_0 \subseteq K' \subseteq K_i$. Then the map $\pi_1((S^3 \setminus \mathcal{W}) \setminus K_i) \to \pi_1((S^3 \setminus \mathcal{W}) \setminus K_0)$ factors through the map $\pi_1((S^3 \setminus \mathcal{W}) \setminus K') \to \pi_1((S^3 \setminus \mathcal{W}) \setminus K_0)$, which is trivial by hypothesis. Rephrasing using the definition of K_i, the map

$$\pi_1(\operatorname{Int} T_i \setminus \mathcal{W}) \to \pi_1(\operatorname{Int} T_0 \setminus \mathcal{W})$$

is trivial. We will show that this leads to a contradiction.

Consider a meridian $\mu_j \subseteq \partial T_j$, for some $j \geq 0$. We claim that μ_j is essential in $T_j \setminus \mathcal{W}$. Suppose it is not. Then, by compactness there is some $k > j$ such that μ_j bounds an immersed disc in the 3-manifold $T_j \setminus T_k$. By Dehn's lemma for 3-manifolds (see e.g. [**Hem76**, Chapter 4]), this implies that the meridian bounds an embedded disc in $T_j \setminus T_k$. Removing a thickening of such an embedded disc from T_j leaves behind a 3-ball containing T_k. This in turn implies that some iterated Whitehead link in S^3 is a split link and since the components of an iterated

Whitehead link are unknotted, that it is the unlink. This is known to be false, and can be seen using, for example, Milnor's invariants of links [**MY11**].

By an abuse of notation, let μ_i also denote a pushed in copy of the meridian of ∂T_i in $\operatorname{Int} T_i$. By the previous paragraph, μ_i is nontrivial in $\pi_1(\operatorname{Int} T_i \smallsetminus \mathcal{W})$. We assert that μ_i is also nontrivial in $\pi_1(\operatorname{Int} T_0 \smallsetminus \mathcal{W})$ (this is another abuse of notation since μ_i gives only a free homotopy class). This is immediate since μ_0 is a product of conjugates of μ_i and thus if μ_i were trivial in $\pi_1(\operatorname{Int} T_0 \smallsetminus \mathcal{W})$, then μ_0 would be trivial in $\pi_1(\operatorname{Int} T_0 \smallsetminus \mathcal{W})$, which would contradict the previous paragraph.

We have found an element of $\pi_1(\operatorname{Int} T_i \smallsetminus \mathcal{W})$ that is mapped to a nontrivial element of $\pi_1(\operatorname{Int} T_0 \smallsetminus \mathcal{W})$, establishing that the map $\pi_1(\operatorname{Int} T_i \smallsetminus \mathcal{W}) \to \pi_1(\operatorname{Int} T_0 \smallsetminus \mathcal{W})$ is nontrivial, as desired. □

The space $S^3 \smallsetminus \mathcal{W}$ is usually called the *Whitehead manifold* and is historically significant. In [**Whi34**], Whitehead attempted to prove the 3-dimensional Poincaré conjecture by claiming that any contractible open 3-manifold is homeomorphic to $\mathbb{R}^3$. The Poincaré conjecture would follow since then any homotopy 3-sphere could be identified with the one-point compactification of the complement of any point, where such a complement would be known to be $\mathbb{R}^3$. Whitehead discovered his error in [**Whi35**] where he showed that the open contractible 3-manifold $S^3 \smallsetminus \mathcal{W}$ is not simply connected at infinity and thus is not homeomorphic to $\mathbb{R}^3$.

7.2. The space $S^3/\mathcal{W}$ is a manifold factor

We consider the decomposition $\mathcal{W} \times \mathbb{R}$ of $S^3 \times \mathbb{R}$, where the decomposition elements of $\mathcal{W} \times \mathbb{R}$ are defined to be the sets $\mathcal{W} \times \{t\}$, for each $t \in \mathbb{R}$ (see also Proposition 4.11). The following is the main result in this chapter.

Theorem 7.5 ([**AR65**]). *The decomposition $\mathcal{W} \times \mathbb{R}$ of $S^3 \times \mathbb{R}$ is shrinkable.*

Remark 7.6. A *manifold factor* is a topological space M such that there exists some other topological space N for which the product $M \times N$ is a manifold. In some sources, the definition requires that $N = \mathbb{R}^n$ for some n. By Theorem 7.5 and Theorem 7.7 below, the space $(S^3 \times \mathbb{R})/(\mathcal{W} \times \mathbb{R})$ is homeomorphic to $S^3 \times \mathbb{R}$. Since $(S^3 \times \mathbb{R})/(\mathcal{W} \times \mathbb{R})$ is homeomorphic to $(S^3/\mathcal{W}) \times \mathbb{R}$ by Proposition 4.11, we see that $S^3/\mathcal{W}$ is a manifold factor with $N = \mathbb{R}$, that is, $S^3/\mathcal{W} \times \mathbb{R}$ is a manifold. Prior to Freedman's proof of the disc embedding theorem, it was already known that the Whitehead link and Whitehead doubles were intimately connected with Casson handles (as we discuss in Chapter 13). The fact that the Whitehead decomposition space $S^3/\mathcal{W}$ is "nearly" a manifold was very encouraging for the attempt to prove the 4-dimensional topological Poincaré conjecture, since it indicated that the additional fourth dimension might make some of the wildness of Casson handles disappear. Hence Theorem 7.5 was historically quite important for the proof of the disc embedding theorem.

So far we have restricted ourselves to compact metric spaces. To prove Theorem 7.5, we will need a more general shrinkability result, which we state below.

Theorem 7.7 ([**Dav07**, Theorem 5.4, p. 26]). *Suppose $\mathcal{D}$ is an upper semicontinuous decomposition of a locally compact, separable metric space X. Then $\pi\colon X \to X/\mathcal{D}$ is approximable by homeomorphisms* (*in the compact open topology*) *if and only if for each compact subset $C \subseteq X/\mathcal{D}$ and each $\varepsilon > 0$, there exists a homeomorphism $h\colon X \to X$ satisfying*

(i) *for all $x \in \pi^{-1}(C) \cup \big(h^{-1} \circ \pi^{-1}(C)\big)$ we have that $d_{X/\mathcal{D}}\big(\pi(x), \pi \circ h(x)\big) < \varepsilon$ and*

(ii) *for each $c \in C$ we have $\operatorname{diam}_X h(\pi^{-1}(c)) < \varepsilon$,*

where $d_{X/\mathcal{D}}$ is some chosen metric on $X/\mathcal{D}$.

Above we have used the fact that the quotient of a potentially noncompact metric space by an upper semi-continuous decomposition is metrisable (Remark 4.14).

Theorem 7.7 is proved by applying Theorem 4.18 to the one-point compactification of the given locally compact, separable space X. The following is a direct corollary.

PROPOSITION 7.8 ([**Dav07**, Theorem 10.1]). *Let M be a d–dimensional manifold equipped with a metric. Let $\mathcal{D}$ be an upper semi-continuous decomposition of M associated with a defining sequence $\{C_i\}_{i=1}^{\infty}$. If for every $\varepsilon > 0$ and every $i \geq 0$ there exists a self-homeomorphism h of $M \times \mathbb{R}$ satisfying*

(a) *the support of h is in $C_i \times \mathbb{R}$,*
(b) *$h(C_i \times \{t\}) \subseteq C_i \times [t-\varepsilon, t+\varepsilon]$ for all $t \in \mathbb{R}$, and*
(c) *$\operatorname{diam}_M h(\Delta \times \{t\}) < 3\varepsilon$ for all decomposition elements Δ of $\mathcal{D}$ and $t \in \mathbb{R}$,*

then the decomposition $\mathcal{D} \times \mathbb{R}$ of $M \times \mathbb{R}$ is shrinkable.

Theorem 7.5 will follow as a special case of the following result.

THEOREM 7.9. *Suppose $\mathcal{D}$ is an upper semi-continuous decomposition of a 3-manifold M associated with a defining sequence $\{C_i\}_{i=1}^{\infty}$, where each connected component of each C_i, for $i \geq 1$, is a solid torus that is null-homotopic in C_{i-1}. Then $\mathcal{D} \times \mathbb{R}$ is a shrinkable decomposition of $M \times \mathbb{R}$.*

The statement of Theorem 7.9 and our proof here follows that of [**AR65**] (see [**Dav07**, Theorem 10.3]) very closely. We will need the following lemma.

LEMMA 7.10. *Let D^2 denote the closed 2-dimensional disc with unit radius and for each $r < 1$ let $D_r \subseteq D^2$ denote the closed subdisc of radius r. Suppose that in the solid torus $T = D^2 \times S^1$, the sub-solid torus $D_r \times S^1$ for some r contains a compact set Q such that the inclusion $Q \to T$ is null-homotopic. Then there exists an $\alpha > 0$ and a self-homeomorphism f of $T \times \mathbb{R}$ such that*

(a) *f is the identity map on $\partial T \times \mathbb{R}$,*
(b) *$f(T \times \{t\}) \subseteq T \times [t-\alpha, t+\alpha]$ for all $t \in \mathbb{R}$, and*
(c) *for each $t \in S^1$, there is a point $s(t) \in S^1$ such that $f(Q \times \{t\}) \subseteq D_r \times \{s(t)\} \times [t-\alpha, t+\alpha]$ for all $t \in \mathbb{R}$.*

Moreover α can be chosen to be arbitrarily small.

PROOF. Let $p\colon D^2 \times \mathbb{R} \to D^2 \times S^1$ be the universal covering map given by $p(b,t) = (b, e^{it})$. Since the inclusion $Q \to T = D^2 \times S^1$ is null-homotopic, we can lift to an embedding $J\colon Q \to D^2 \times \mathbb{R}$. Let $w\colon Q \to \mathbb{R}$ be the map J composed with the projection $D^2 \times \mathbb{R} \to \mathbb{R}$. Observe that for each $(b,s) \in Q$, we have

$$e^{iw(b,s)} = s. \tag{7.1}$$

Since Q is compact, $w(Q) \subseteq [-\alpha, \alpha]$ for some $\alpha \in \mathbb{R}$. By the Tietze extension theorem there is a $\psi\colon T \to [-\alpha, \alpha]$ extending w such that $\psi(\partial T) = 0$.

Next we define two self-homeomorphisms λ and τ of $T \times \mathbb{R}$; the desired homeomorphism f will be given as $\tau \circ \lambda$. For $(u,t) \in T \times \mathbb{R}$ the translation λ is defined to be

$$\lambda(u,t) := (u, \psi(u) + t).$$

The rotation τ is defined as $\tau(u,t) := (\tau_t(u), t)$ using a self-homeomorphism τ_t of the solid torus T, where for any $(b,s) \in D^2 \times S^1 = T$

$$\tau_t(b,s) := (b, se^{-it})$$

if $|b| \leq r$ and

$$\tau_t(b,s) = (b, se^{-it\frac{1-|b|}{1-r}})$$

if $|b| > r$. The maps τ and λ are depicted in Figure 7.2.

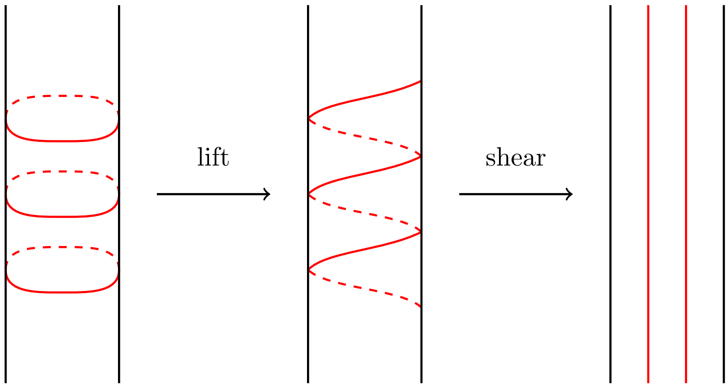


Figure 7.2. The homeomorphism λ followed by the rotation/shear τ.

Now consider the map $f := \tau \circ \lambda$. By definition, both λ and τ act as the identity on $\partial T \times \mathbb{R}$, which implies (a). Moreover λ modifies the second coordinate t by at most α and τ preserves the second coordinate, which implies (b). To show (c) let $\pi\colon T \times \mathbb{R} \to T$ be the projection onto T. Then we see that for $(b, s) \in Q$

$$\pi \circ f(b, s, t) = (b, se^{-i(\psi(b,s)+t)}) = (b, se^{-i(w(b,s)+t)}),$$

which is equal to (b, e^{-it}) by (7.1), which together with (b) implies (c).

To see that α can be chosen to be arbitrarily small, replace e^{ix} by $e^{ix/\varepsilon}$ in each previous formula, where $\varepsilon > 0$ is arbitrarily small. This means that we reparameterise the covering map p and hence $\psi(T) \subseteq [-\varepsilon\alpha, \varepsilon\alpha]$. □

Proof of Theorem 7.9. Choose an $\varepsilon > 0$ and an integer $i \geq 1$. We wish to construct a self-homeomorphism h of $M \times \mathbb{R}$ satisfying the requirements of Proposition 7.8.

Let T_1 be a component of C_i. We identify T_1 with the solid torus $D^2 \times S^1$. Let $\widetilde{h}_1$ be a self-homeomorphism of T_1 that is the identity on ∂T_1 and which performs a "radial shrinking" in T_1, that is, for all $s \in S^1$ the diameter of $\widetilde{h}_1(D_r \times \{s\}) < \varepsilon$.

Let Q denote $T_1 \cap C_{i+1}$. By assumption, the embedding $Q \to D^2 \times S^1$ is nullhomotopic and lies in $D_r \times S^1$, where D_r is the subdisc of the unit disc D^2 of some radius $r < 1$. Apply Lemma 7.10 to find $0 < \alpha < \varepsilon$ and a self-homeomorphism f of $D^2 \times S^1 \times \mathbb{R}$ satisfying (a), (b), and (c) of Lemma 7.10. Then $f \circ (\widetilde{h}_1 \times \mathrm{id}_{\mathbb{R}})$ is a self-homeomorphism of $T_1 \times \mathbb{R}$. Since this map acts by the identity on $\partial T_1 \times \mathbb{R}$, it extends via the identity to a self-homeomorphism h_1 of $M \times \mathbb{R}$.

Repeat this construction for all other components of C_i to obtain the self-homeomorphisms $\{h_i\}_{i=0}^n$ of $M \times \mathbb{R}$. The desired self-homeomorphism h of $M \times \mathbb{R}$ is defined to be $h_n \circ \cdots \circ h_1$. This h satisfies all the requirements of Proposition 7.8 and so the decomposition $\mathcal{D} \times \mathbb{R}$ of $M \times \mathbb{R}$ is shrinkable. This finishes the proof of Theorem 7.9. □

CHAPTER 8

Mixed Bing-Whitehead decompositions

Daniel Kasprowski and Min Hoon Kim

So far we have considered decompositions of S^3 obtained by defining patterns coming from either Bing or Whitehead doubling the core of a solid torus. In this chapter, we consider *mixed Bing-Whitehead decompositions*. These decompositions arise from a defining sequence of nested solid tori where each element is either a (thickened) Bing double or Whitehead double of the cores of the solid tori in the previous stage. The Bing decomposition and the Whitehead decomposition are then seen to be special degenerate instances of mixed Bing-Whitehead decompositions. Since the Bing decomposition shrinks and the Whitehead decomposition does not, it is perhaps not surprising that the more Bing doubling stages there are in a mixed Bing-Whitehead decomposition, the more likely it is to shrink. The goal of this chapter is to make this statement precise for mixed Bing-Whitehead decompositions. We will also consider slight generalisations, called *mixed ramified Bing-Whitehead decompositions*. The shrinking of these decompositions is addressed in Scholium 8.13, which is a key ingredient in our proof of the disc embedding theorem in Part IV.

A *mixed Bing-Whitehead decomposition* is defined to be a decomposition of S^3 of the form

$$\mathcal{B}^{b_1}\ \mathcal{W}\ \mathcal{B}^{b_2}\ \mathcal{W}\ \mathcal{B}^{b_3}\ \cdots, \tag{8.1}$$

where $\mathcal{B}$ indicates Bing doubling and $\mathcal{W}$ indicates Whitehead doubling. More precisely, we have the following defining sequence. Start with the standard unknotted solid torus T_0 in S^3. Then for $i = 1, \ldots, b_1$, let T_i denote the collection of solid tori obtained as a regular neighbourhood of the (untwisted) Bing double of the cores of the solid tori which form T_{i-1}. In other words, we use the defining pattern given in Figure 5.7. Then T_{b_1+1} is a regular neighbourhood of the (untwisted) Whitehead double of the cores of the solid tori which form T_{b_1}. We could use the defining pattern given in Figure 7.1 or the pattern with the opposite clasp. The clasps need not match across different components. Then we construct the solid tori forming T_i for $i = b_1+2, \ldots, b_1+b_2+1$ by Bing doubling as before, and the solid tori forming $T_{b_1+b_2+2}$ by Whitehead doubling, and so forth. As usual, the decomposition consists of the connected components of the infinite intersection $\bigcap_{i=0}^{\infty} T_i$.

We permit that $b_j = 0$ for some j, or for all j, in which case we get the Whitehead decomposition (for some choice of clasps). However, in our later applications this will never be the case. Indeed, we will be specifically interested in the case where $b_j \geq 4$ for all j.

We also need to consider a slightly more general construction, involving *ramification*. That is, instead of applying Bing or Whitehead doubling to the cores of the previous set of solid tori, we instead apply Bing or Whitehead doubling to multiple unlinked parallel copies of the cores. The number of parallel copies taken need not match up across different connected components. The resulting decompositions are called *mixed ramified Bing-Whitehead decompositions*. We will see that our arguments for shrinking are independent of ramification, since we construct homeomorphisms that are fixed outside certain solid tori. For ease of notation, we will ignore ramifications and choices of clasps whenever possible.

Ancel and Starbird [**AS89**], as well as [**Wri89**], determined exactly when mixed Bing-Whitehead decompositions shrink.

THEOREM 8.1 ([**AS89**, **Wri89**]). *A mixed Bing-Whitehead decomposition as defined in* (8.1) *shrinks if and only if the series* $\sum_{i=1}^{\infty} \frac{b_i}{2^i}$ *diverges.*

The proof combines the two techniques highlighted in Chapters 5 and 6, namely Bing's rotation trick to show that the Bing decomposition shrinks (Theorem 5.4) and the tracking of intersections with certain meridional discs to show that the decomposition $\mathcal{B}_2$ does not shrink (Theorem 6.1). Heuristically speaking, each Whitehead double sets us back in terms of shrinking, but we can make up for it with enough layers of Bing doubling.

Kasprowski and Powell [**KP14**] studied when decompositions given by defining sequences consisting of embedded solid tori, called *toroidal decompositions*, shrink. We define these more precisely below. Their result builds on [**AS89**, **Wri89**] and gives an effective way to determine if a toroidal decomposition is shrinkable. In this chapter we present the results of [**KP14**], since it is useful to understand the situation of mixed Bing-Whitehead decompositions as a special case of a more general phenomenon.

8.1. Toroidal decompositions

Let $\mathcal{L} = \{L^i\}_{i=1}^{\infty}$ be a sequence of oriented links in S^3, where each

$$L^i = L_0^i \sqcup L_1^i \sqcup \cdots \sqcup L_{n_i}^i$$

is an (n_i+1)-component link in S^3 with a specified component L_0^i unknotted. Each link L^i determines an embedding

$$\widehat{L}^i := L_1^i \sqcup \cdots \sqcup L_{n_i}^i \colon \bigsqcup_{j=1}^{n_i} S^1 \times D^2 \to S^1 \times D^2,$$

where the target $S^1 \times D^2$ is the closure of the complement of a regular neighbourhood of L_0^i, which is identified with $S^1 \times D^2$ using the untwisted framing of L_0^i, and the map consists of taking regular neighbourhoods of the components $L_1^i \sqcup \cdots \sqcup L_{n_i}^i$. Let $I_0 := 1$ and $I_s := \prod_{i=1}^{s} n_i$ for $s \geq 1$. (Recall that each link L^i has $n_i + 1$ components.) Using the embeddings $\{\widehat{L}^s\}$ and integers $\{I_s\}$, we define a sequence of nested solid tori

$$T_0 \supsetneq T_1 \supsetneq T_2 \supsetneq \cdots$$

as follows. Define T_0 to be the standard unknotted solid torus in S^3. For $s \geq 1$, we define T_s inductively as the image of the embedding

$$\bigsqcup_{I_{s-1}} \widehat{L}^s \colon \bigsqcup_{I_{s-1}} \bigsqcup_{k=1}^{n_s} S^1 \times D^2 \to \bigsqcup_{I_{s-1}} S^1 \times D^2 = T_{s-1}.$$

That is, we use the same link for each component of each stage, but potentially different links for different stages.

The sequence $T_0 \supsetneq T_1 \supsetneq T_2 \supsetneq \cdots$ of nested solid tori is a defining sequence for the decomposition $\mathcal{D}$ of S^3 consisting of the connected components of $\bigcap_{s=0}^{\infty} T_s$, called the *toroidal decomposition of S^3 associated to $\mathcal{L}$*. Toroidal decompositions are easily seen to be upper semi-continuous using Proposition 4.10.

EXAMPLE 8.2. Suppose that $\mathcal{L} = \{L^i\}$, where each L^i is either a Whitehead link or the Borromean rings. Then $\widehat{L}^i$ is either a Whitehead double or the Bing double of the core $S^1 \times \{0\}$ of $S^1 \times D^2$, and the toroidal decomposition $\mathcal{D}$ associated to $\mathcal{L}$ is of the form (8.1), where possibly some $b_j = 0$. When L^i is the Whitehead link for all i, we get the Whitehead decomposition, whereas when L^i is the Borromean rings for all i, we get the Bing decomposition.

REMARK 8.3. Since the link used in each component of each stage is constant, toroidal decompositions include mixed ramified Bing-Whitehead decompositions, where the ramification and clasps are constant within each stage of the defining sequence. One can expand the definition to allow different ramifications and clasp signs within each stage. The arguments of this chapter will still apply. As discussed above, we shall ignore ramifications and clasp signs as much as possible in this chapter, since the disc replicating functions and the veracity of Theorem 8.4 below do not depend on these modifications.

8.2. Disc replicating functions

Next we discuss the shrinkability criterion for toroidal decompositions given in [**KP14**].

THEOREM 8.4 ([**KP14**]). *Let $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. For every link $L \subseteq S^3$ with a specified unknotted component there is a function $D_L \colon \mathbb{N}_0 \to \mathbb{N}_0$ with $D_L(0) = 0$, called the* disc replicating function*, which is well defined up to isotopy of L.*

The toroidal decomposition $\mathcal{D}$ associated to $\mathcal{L} = \{L^i\}$ is shrinkable if and only if for all $k \geq 0$ and $m \geq 1$,

$$\lim_{p\to\infty} (D_{L^{m+p}} \circ \cdots \circ D_{L^m}(k)) = 0.$$

It follows that the shrinkability of $\mathcal{D}$ depends only on the isotopy classes of the links $\{L^i\}$ (compare with Proposition 4.7). To define the disc replicating function D_L we will need the following notions from [**Wri89**].

DEFINITION 8.5 (k-interlacing of meridional discs). Two collections of pairwise disjoint meridional discs $A = \bigsqcup_{i=1}^k A_i$, $B = \bigsqcup_{i=1}^k B_i$ of $S^1 \times D^2$ are said to form a *k-interlacing of meridional discs* if each component of $S^1 \times D^2 \smallsetminus (A \cup B)$ has precisely one A_i and one B_i in its closure.

DEFINITION 8.6 (Meridional k-interlacing). A *meridional k-interlacing* is a collection of pairwise disjoint meridional discs $A = \bigsqcup A_i$ and $B = \bigsqcup B_i$ of $S^1 \times D^2$ such that there are subsets $A' \subseteq A$ and $B' \subseteq B$ forming a k-interlacing of meridional discs, but no such subsets forming a $k+1$-interlacing of meridional discs. If there are no such subsets of A and B, then we say that A and B form a meridional 0-interlacing.

Let L be a link with a specified unknotted component L_0. As above, let $\widehat{L}$ denote the corresponding link $L \smallsetminus L_0$ in $S^1 \times D^2 = \overline{S^3 \smallsetminus \nu(L_0)}$, identified using the untwisted framing on L_0. For a meridional k-interlacing (A, B) of a solid torus $S^1 \times D^2$, we define $D_L(A, B) \in \mathbb{N}_0$ to be the minimal integer ℓ such that there exists a link L' in $S^1 \times D^2$ ambiently isotopic to $\widehat{L}$ in $S^1 \times D^2$ and a closed regular neighbourhood $\nu(L')$ whose boundary intersects both A and B transversely and only in meridians of $\nu(L')$, and for which all components of $\nu(L')$ inherit at most a meridional ℓ-interlacing by intersections with $A \cup B$. We have the following lemma, whose proof we omit.

LEMMA 8.7 ([**KP14**, Lemma 2.8]). *For every two meridional k-interlacings (A, B) and (A', B'), and each link $L \subseteq S^3$, $D_L(A, B)$ and $D_L(A', B')$ agree.*

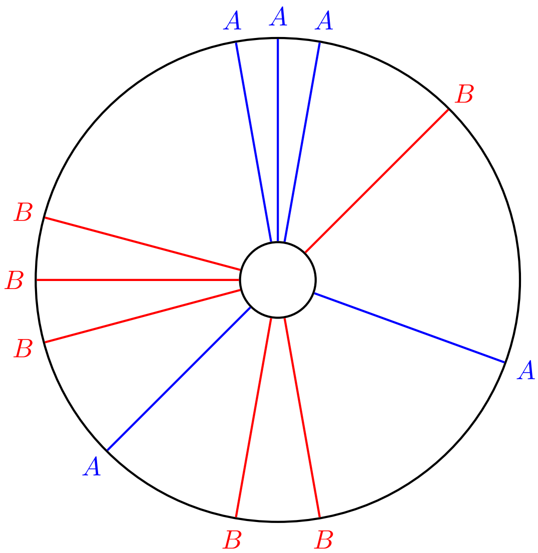


Figure 8.1. A meridional 3-interlacing.

As a consequence, the quantity $D_L(A, B)$ does not depend on the choice of a meridional k-interlacing (A, B) and we define $D_L(k) := D_L(A, B) \in \mathbb{N}_0$, where (A, B) is some choice of k-interlacing.

We also observe here for future use that the disc replicating function is non-decreasing for every link. That is, for every link L, if $i < j$, then $D_L(i) \leq D_L(j)$. This is a direct consequence of the definition.

8.3. Shrinking of toroidal decompositions

Now we begin to address the shrinkability of toroidal decompositions. The following lemma describes what we need to prove.

LEMMA 8.8. *Let $\mathcal{L}$ be a sequence of links with associated toroidal decomposition $\mathcal{D}$ and defining sequence $\{T_i\}$ consisting of nested solid tori.*

The decomposition $\mathcal{D}$ shrinks if for every $m \geq 1$ and for every $\varepsilon > 0$, there exist $p \geq 1$ and a homeomorphism $h\colon S^1 \times D^2 \to S^1 \times D^2$ that is the identity outside T_m and such that for each component $(T_{m+p})_i$ of T_{m+p}, we have that $\operatorname{diam} h((T_{m+p})_i) < \varepsilon$.

PROOF. We will directly use Definition 4.16. Let $\pi\colon S^1 \times D^2 \to S^1 \times D^2/\mathcal{D}$ be the quotient map. Fix $\varepsilon > 0$. The space $X/\mathcal{D}$ is metrisable by Corollary 4.13 since $\mathcal{D}$ is upper semi-continuous. Fix a metric on $X/\mathcal{D}$

CLAIM. *For every $\varepsilon > 0$ there exists $q \geq 1$ such that for each torus T' in T_q we have* $\operatorname{diam} \pi(T') < \varepsilon$.

Next we finish the proof assuming the claim. By the claim, there exists an integer $m \geq 1$ such that $\operatorname{diam} \pi(T') < \varepsilon$ for every component T' of T_m. By the hypothesis of the lemma, for every m, and therefore for this m, there exists $p \geq 1$ and a homeomorphism $h\colon S^1 \times D^2 \to S^1 \times D^2$ that is the identity outside T_m, such that for each component $(T_{m+p})_i$ of T_{m+p}, we have $\operatorname{diam} h((T_{m+p})_i) < \varepsilon$. Since $\operatorname{diam} h((T_{m+p})_i) < \varepsilon$, the second condition of Definition 4.16 holds. To check the first condition, let $x \in S^1 \times D^2$. If x and $h(x)$ are different, then x is contained inside T_m and $h(x)$ has to lie in the same component T' of T_m, since h is the identity outside T_m by hypothesis. Therefore, by our choice of m, $d(\pi \circ h(x), \pi(x)) < \varepsilon$. That is, the first condition of Definition 4.16 also holds and therefore $\mathcal{D}$ shrinks.

It remains to prove the claim. Suppose the statement is not true. Then there exists an infinite sequence of solid tori $\{T_i'\}$ in $S^1 \times D^2$ with $\operatorname{diam} \pi(T_i') > \varepsilon$. We can assume $\{T_i'\}$ is a nested sequence. As the intersection of the nested, connected, compact sets $\{T_i'\}$, $I := \bigcap_{i=n}^{\infty} T_i'$ is nonempty and connected by Lemma 7.1. In each T_i', choose two points x_i and y_i with $d(\pi(x_i), \pi(y_i)) \geq \varepsilon$. After passing to subsequences, we can assume that $\{x_i\}$ and $\{y_i\}$ converge in $S^1 \times D^2$ since $S^1 \times D^2$ is (sequentially) compact. Let x and y be the limits of the sequences $\{x_i\}$ and $\{y_i\}$ respectively. It is easy to check that $x, y \in I$. Additionally, $\pi(I)$ is a point since I is a decomposition element and thus $\pi(x) = \pi(y)$. On the other hand we will show that $d(\pi(x), \pi(y)) \geq \frac{\varepsilon}{2}$, which will yield the desired contradiction.

Note that there exists N such that $d(\pi(x_i), \pi(x)) \leq \frac{\varepsilon}{4}$ and $d(\pi(y_i), \pi(y)) \leq \frac{\varepsilon}{4}$ for every $i \geq N$, since the sequences $\{x_i\}$ and $\{y_i\}$ converge by hypothesis and π is continuous. Then for every $i \geq N$ we have

$$
\begin{aligned}
\varepsilon \leq d(\pi(x_i), \pi(y_i)) &\leq d(\pi(x_i), \pi(x)) + d(\pi(x), \pi(y)) + d(\pi(y_i), \pi(y)) \\
&\leq \frac{\varepsilon}{4} + d(\pi(x), \pi(y)) + \frac{\varepsilon}{4}.
\end{aligned}
$$

Thus $d(\pi(x), \pi(y)) \geq \frac{\varepsilon}{2}$ as desired, which completes the proof of the claim. □

Proof of Theorem 8.4. We only give the proof that a toroidal decomposition $\mathcal{D}$ given by the sequence of links $\mathcal{L} = \{L^i\}$ shrinks if for all $k \geq 0$ and $m \geq 1$, we have that

$$\lim_{p \to \infty} (D_{L^{m+p}} \circ \cdots \circ D_{L^m}(k)) = 0.$$

This is the direction we will need for the proof of the disc embedding theorem. For the other direction, see [**KP14**]. The argument is an extension of the proof in Chapter 6. We indicated the proof that D_L is well defined in Section 8.2.

Our goal is to apply Lemma 8.8. Given $\varepsilon > 0$ and $m \geq 1$, choose a $k \geq 1$ and k-interlacing (A, B) of each component of T_{m-1} such that each component of $T_{m-1} \smallsetminus (A \cup B)$ has diameter strictly less than $\varepsilon/2$. By the definition of the disc replicating function, we may isotope T_m within T_{m-1} so that each component of T_m intersects $A \cup B$ in at most a meridional $D_{L^m}(k)$-interlacing.

Now we iterate. Let T be a component of T_m which inherited a meridional $D_{L^m}(k)$-interlacing in the previous step. Then by the definition of the disc replicating function, there is an isotopy of T_{m+1} within T so that the components meet $A \cup B$ in at most a meridional $D_{L^{m+1}}(D_{L^m}(k))$-interlacing. Apply this isotopy, extended via the identity to all of S^3. Now suppose T is a component of T_m which inherited a meridional ℓ-interlacing in the previous step where $\ell < D_{L^m}(k)$. Again by the definition of the disc replicating function, there is an isotopy of T_{m+1} within T so that the components meet $A \cup B$ is at most a meridional $D_{L^{m+1}}(\ell)$-interlacing. Apply this isotopy within T, extended via the identity to all of S^3. Note that since the disc replicating function is non-decreasing, $D_{L^{m+1}}(\ell) \leq D_{L^{m+1}}(D_{L^m}(k))$, so at the end of this iteration, we have that each component of T_{m+1} in T_m intersects $A \cup B$ in at most a meridional $D_{L^{m+1}} \circ D_{L^m}(k)$-interlacing.

Continue this process: at each step, apply small isotopies within solid tori as dictated by the disc replicating function. Since the disc replicating function takes non-negative integer values and

$$\lim_{p \to \infty} (D_{L^{m+p}} \circ \cdots \circ D_{L^m}(k)) = 0$$

by hypothesis, there exists some $p \geq 1$ such that $(D_{L^{m+p}} \circ \cdots \circ D_{L^m}(k)) = 0$. Thus, after p iterations, we will have arranged that each component of T_{m+p} intersects $A \cup B$ in a meridional $(D_{L^{m+p}} \circ \cdots \circ D_{L^m}(k))$-interlacing i.e. a meridional 0-interlacing.

Recall that by definition, this means each component of T_{m+p} meets at most one component of A or one component of B. Since we arranged at the start that each component of $T_{m-1} \smallsetminus (A \cup B)$ has diameter strictly less than $\varepsilon/2$, it follows that the diameter of each component of T_{m+p} is strictly less than ε. The desired homeomorphism h is defined to be the result of the composition of all of the isotopies applied in the iterative process. This completes the proof by Lemma 8.8. $\square$

8.4. Computing the disc replicating function

The results above show that the disc replicating function gives a precise way to determine when a toroidal decomposition shrinks. However, we have said nothing so far about how to compute the disc replicating function for a given toroidal decomposition. In [**KP14**], the disc replicating function was computed precisely for a certain family of toroidal decompositions, including the mixed Bing-Whitehead decompositions which are of particular relevance to us. This is accomplished using a computable lower bound, which we define next.

Fix some link $L = L_0 \sqcup \cdots \sqcup L_m$ with the specified unknotted component L_0 and let $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. We will presently define a function $f_L \colon \mathbb{N} \cup \{0\} \to \mathbb{N} \cup \{0\}$, which is interesting due to the following theorem.

Theorem 8.9 ([**KP14**, Theorem 4.5]). *For every $k \in \mathbb{N}_0$ and link $L \subseteq S^3$, $f_L(k) \leq D_L(k)$.*

Let J be a link obtained from L as a result of applying the following operations.

(1) Take the d-fold cover of S^3 branched along L_0 for some $d \geq 1$. The resulting 3-manifold is again S^3. Let $\widetilde{J}$ and $\widehat{J}_0$ be the pre-image of L and L_0 respectively.
(2) Take a sublink $\widehat{J}$ of $\widetilde{J}$ that includes the unknotted component of $\widetilde{J}$, which is now called $\widehat{J}_0$.
(3) Perform ± 1-framed Dehn surgery along ℓ unknotted components of $\widehat{J} \smallsetminus \widehat{J}_0$, for some $\ell \geq 0$, lying in an open ball in $S^3 \smallsetminus \nu(\widehat{J}_0)$. The resulting 3-manifold is a copy of S^3. The image therein of the remaining components of $\widehat{J}$ forms the link J, with the specified unknotted component J_0 which is the image of $\widehat{J}_0$.

We will now define a correspondence $f_L^J \colon \mathbb{N}_0 \to \mathbb{N}_0$ by the following rules. Recall that given an ordered, oriented n-component link K and a multi-index I, consisting of a possibly repeating list of k elements of $\{1, \dots, n\}$ for some $k \geq 2$, we have the Milnor invariant $\overline{\mu}_I(K)$, which is a link isotopy invariant [**Mil57**]. In particular, the length 2 Milnor invariant $\overline{\mu}_{ij}(K)$ is the linking number of the ith and jth components of L, for $i \neq j$.

(1) Define $f_L^J(0) = 0$.
(2) If there is no multi-index I containing zero such that $\overline{\mu}_I(J) \neq 0$, then define $f_L^J(k) = 0$ for all k.
(3) Otherwise, define $f_L^J(k)$ to the maximal non-negative integer that can be obtained by one of the following procedures.
    (i) Suppose there exists a multi-index I containing zero and with length $|I| = 2$ such that $\overline{\mu}_I(J) \neq 0$, that is there is some component of J with nonzero linking number with J_0. Then let J^* be a component of J with maximal linking number with J_0, and set
$$f_L^J(k) := |\ell\mathrm{k}(J^*, J_0)| \cdot k$$
for all $k \geq 1$.

(ii) If there exists a multi-index I containing zero and $|I| > 2$ such that $\overline{\mu}_I(J) \neq 0$, then define
$$f_L^J(k) = \left\lceil \frac{2dk}{n+\ell} \right\rceil - 1,$$
where n is the number of components of $J \smallsetminus J_0$. Recall that d is the index of the branched cover constructed and ℓ is the number of times ± 1-framed Dehn surgery was performed in the construction of J.

Now to obtain an invariant of L we take another maximum.

Definition 8.10. For each $k \in \mathbb{N}_0$, let $f_L(k)$ be the maximum value of $f_L^J(k)$ for the links J produced above.

It might be difficult to determine $f_L(k)$ precisely in general. However, in practice one just needs to determine sufficiently good lower bounds.

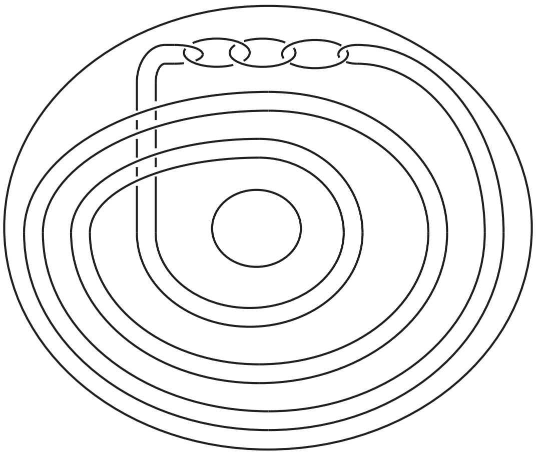


Figure 8.2. An (n, m) link for $n = 4$, $m = 3$, shown in the complementary solid torus for the unknotted component.

Example 8.11. An (n, m)-link, for $n, m \geq 1$ is an $(n+1)$-component link produced as follows. A $(1, 1)$ link is a Whitehead link. An $(n, 1)$ link with $n \geq 2$ has a specified unknotted component as well as a chain of n components, each linked with the previous component and the next with linking number ± 1, in a circular pattern in the complement of the unknotted component. Thus, the Borromean rings is a $(2, 1)$-link. In an (n, m)-link for $m > 1$ the components in a chain wrap around the solid torus defined by the unknotted component, as shown in Figure 8.2. No further weaving of the chain is allowed. Note that the defining pattern of the decomposition $\mathcal{B}_2$ from Chapter 6 and the meridian of the solid torus containing it (see Figure 6.1) form a $(2, 2)$-link.

Proposition 8.12 ([**KP14**, Proposition 5.1]). *Let $L_{n,m}$ be an (n, m)-link for some $n, m \geq 1$. Then*
$$D_L(k) = \max\left\{ \left\lceil \frac{2mk}{n} \right\rceil - 1, 0 \right\},$$
for each $k \in \mathbb{N}_0$.

In particular the Bing decomposition is defined using a $(2, 1)$-link, and $D_{L_{2,1}}(k) = k - 1$ for $k \geq 1$, while the Whitehead decomposition is defined using a $(1, 1)$-link, namely a Whitehead link, and $D_{L_{1,1}}(k) = 2k - 1$ for $k \geq 1$.

The proof of the above result uses Theorem 8.9. In order to compute the function f_L, Kasprowski and Powell construct links for which the Milnor invariants are easily computed. In particular, for (n, m)-links, one can reduce to the case of the Whitehead link or the Borromean rings. An upper bound for the disc replicating function can be constructed directly.

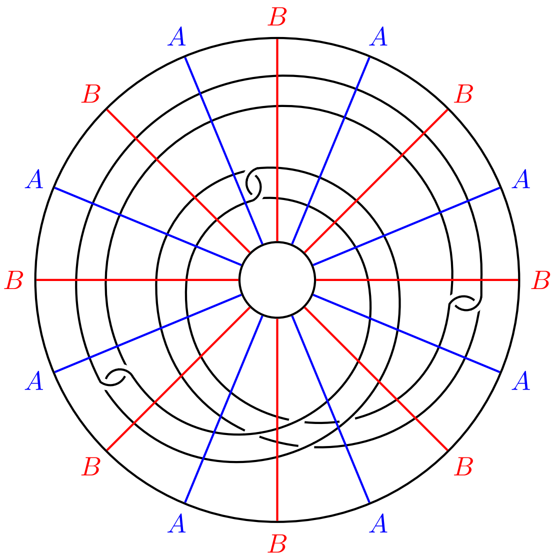


Figure 8.3. A $(3, 2)$ link intersecting an 8-interlacing optimally. Here $D_L(8) = 10$.

Combined with Theorem 8.4, we now have another way to detect the shrinkability (or lack theoreof) of the decompositions $\mathcal{B}$, $\mathcal{B}_2$, and $\mathcal{W}$. Note that $\mathcal{B}$, $\mathcal{B}_2$, and $\mathcal{W}$ are the toroidal decompositions associated to the constant sequences of $(2, 1)$-, $(2, 2)$-, and $(1, 1)$-links respectively. From Proposition 8.12, as mentioned above we see that

$$D_{L_{2,1}}(k) = k - 1,\ D_{L_{2,2}}(k) = 2k - 1,\ \text{and } D_{L_{1,1}}(k) = 2k - 1.$$

By Theorem 8.4, it follows immediately that $\mathcal{B}$ shrinks, but $\mathcal{B}_2$ and $\mathcal{W}$ do not.

More generally, we can consider a sequence $\mathcal{L}$ consisting of (n_i, m_i)-links given by $\{L_{n_i,m_i}\}_{i=1}^{\infty}$. Define $\tau_i := \frac{n_i}{2m_i}$. Let $\mathcal{D}$ be the toroidal decomposition associated to $\mathcal{L}$. Using Theorem 8.4 and Proposition 8.12, we obtain the following criteria to detect the shrinking of the decomposition $\mathcal{D}$ [**KP14**, Corollary 5.2].

(1) If $\sum_{j=1}^{\infty} \prod_{i=1}^{k} \tau_i$ converges, then $\mathcal{D}$ does not shrink.
(2) If $\sum_{j=1}^{\infty} \left(n_j^{-1} \prod_{i=1}^{j} \tau_i \right)$ diverges, then $\mathcal{D}$ shrinks.
(3) If $\sup_{i \in \mathbb{N}} n_i < \infty$, then $\mathcal{D}$ shrinks if and only if $\sum_{j=1}^{\infty} \prod_{i=1}^{k} \tau_i$ diverges.

The final criterion is sufficient to reprove Theorem 8.1, originally due to Ancel and Starbird [**AS89**] and Wright [**Wri89**], as we now show.

Proof of Theorem 8.1. The following proof is given in [**KP14**, Section 5.1]. For a given sequence of non-negative integers b_i, recursively define a sequence of positive integers by setting $w_0 = 1$ and $w_i = w_{i-1} + b_i + 1$ for $i > 1$. Let L^{w_i} be a Whitehead link and let L^j be the Borromean rings for $j \notin \{w_i \mid i \geq 1\}$. Let $\mathcal{L}$ be the sequence of links consisting of the L^i. Note that the decomposition defined as in (8.1) is the decomposition $\mathcal{D}$ associated to $\mathcal{L}$. Let $\tau_{w_i} = 1/2$ and let $\tau_\ell = 1$ for $\ell \notin \{w_i \mid i \geq 1\}$. This concords with a Whitehead link being a $(1, 1)$-link, so

$\tau_{w_i} = \frac{1}{2\cdot 1} = \frac{1}{2}$, and the Borromean rings being a $(2,1)$-link, so $\tau_\ell = \frac{2}{2\cdot 1} = 1$. By the third criterion given above, $\sum_{j=1}^{\infty}\left(\prod_{i=1}^{j}\tau_i\right) = 2\sum_{j=1}^{\infty}\frac{b_j}{2^j}$ diverges if and only if the decomposition $\mathcal{D}$ shrinks, since n_i is either 1 or 2 in this case. $\square$

We observe that the disc replicating function for a ramified decomposition is the same as for the unramified version and thus Theorem 8.1 applies equally well to mixed ramified Bing-Whitehead decompositions. Indeed, with a bit more work, we can establish the following result. Along with Theorem 8.1, this is what we will need from this chapter moving forward.

SCHOLIUM 8.13. *Let $\mathcal{D}_{BW}$ be a mixed ramified Bing-Whitehead decomposition of $D^2\times S^1$ defined by a sequence of nested solid tori as in (8.1) that shrinks, i.e. such that $\sum_j \frac{b_j}{2^j}$ diverges. Then there is a sequence of homeomorphisms $h_i\colon D^2\times S^1 \to D^2\times S^1$, for $i\geq 1$, with each h_i isotopic to the identity via an isotopy supported in the ith stage of the defining sequence for $\mathcal{D}_{BW}$, such that*

$$\lim_{m\to\infty} h_m\circ\cdots\circ h_1\colon D^2\times S^1\to D^2\times S^1$$

exists and has inverse sets coinciding with the quotient map $\pi\colon D^2\times S^1 \to D^2\times S^1/\mathcal{D}_{BW}$.

In particular, the defining sequence of the mixed ramified Bing-Whitehead decomposition can be repositioned so that each element of the new decomposition is a single point.

PROOF. Choose a monotone decreasing sequence $\{\varepsilon_i\}_{i\geq 1}$ with $\varepsilon_i\to 0$ as $i\to\infty$. Let $k_0=0$. For each ε_i, the proofs of Theorem 8.1 and Theorem 8.4 above gives a finite sequence of homeomorphisms $h_{k_{i-1}}, h_{k_{i-1}+1}\ldots, h_{k_i}$, for some k_i, where each h_ℓ is isotopic to the identity and supported in the ℓth stage of the defining sequence for $\mathcal{D}_{BW}$, such that the composition $h_{k_i}\circ\cdots\circ h_1\colon D^2\times S^1\to D^2\times S^1$ shrinks each of the decomposition elements to size less that ε_i. More precisely, go deep in the defining sequence to the $(k_{i-1}+m)$th stage, for some m, so that the image of every $(k_{i-1}+m)$th stage solid torus under $h_{k_{i-1}}\circ\cdots\circ h_1$ has diameter less than ε_i in the chosen metric of $D^2\times S^1/\mathcal{D}_{BW}$, as in the claim in the proof of Lemma 8.8. Define $h_{k_{i-1}+1}=\cdots=h_{k_{i-1}+m}=\mathrm{Id}$. Then choose a meridional interlacing on the solid tori in the $(k_{i-1}+m)$th stage close enough so that the proof of Theorem 8.4 applies to make the decomposition elements of diameter less than ε_i after a sequence of homeomorphisms. This requires a finite number q of isotopies, which give rise to homeomorphisms $h_{k_{i-1}+m+1},\ldots,h_{k_{i-1}+m+q}$ of $D^2\times S^1$. Set $k_i := k_{i-1}+m+q$. Iterate this procedure to yield a sequence of homeomorphisms with the required properties.

For the final sentence, consider the commutative diagram

$$\begin{tikzcd} D^2\times S^1 \arrow[r, "\lim h_m"] \arrow[d, "\pi"] & D^2\times S^1 \\ D^2\times S^1/\mathcal{D}_{BW} \arrow[ur, "\cong"] & \end{tikzcd}$$

where the diagonal arrow is defined and indeed a homeomorphism by the previous paragraph. Since the right-most copy of $D^2\times S^1$ contains the repositioned defining sequence, indeed each element of the (repositioned) mixed ramified Bing-Whitehead decomposition is a singleton set. $\square$

CHAPTER 9

Shrinking starlike sets

Jeffrey Meier, Patrick Orson, and Arunima Ray

So far we have focussed on studying the behaviour of certain specific decompositions. To show that these decompositions shrink, we have produced explicit homeomorphisms. In this chapter, we will work more abstractly. We will prove one of the main shrinking results needed to shrink the complement of "the design in the standard 2-handle" in Part IV; this is called the α *shrink*. The relevant decomposition, consisting of the holes$^+$, is *null* and *recursively starlike-equivalent*. In this chapter, we will define these terms and prove the corresponding shrinking theorem. We will apply the theorem in Part IV by checking the conditions on the decomposition elements and we will not need to construct explicit homeomorphisms. Similarly to Chapter 8, we present the theorem as a special case of a more general phenomenon [**MOR19**, Theorem 1.1], since we believe this most brightly illuminates the key ideas. Comparatively more *ad hoc* arguments were given in both [**Fre82a**] and [**FQ90**].

We introduce some notation for this chapter. Let (X, d) be a metric space. For a subset $A \subseteq X$, any $x_0 \in X$, and $r > 0$, write

$$N(A; r) := \{x \in X \mid d(x, A) < r\}$$

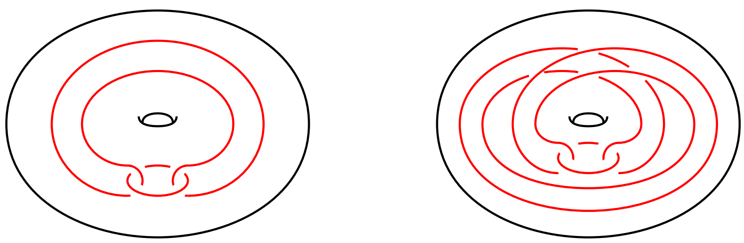
and

$$B(x_0; r) := \{x \in X \mid d(x, x_0) < r\}.$$

9.1. Null collections and starlike sets

Definition 9.1. A collection of subsets $\{T_i\}_{i \in I}$ of a metric space X is said to be *null* if for every $\varepsilon > 0$ there are only finitely many $i \in I$ such that $\operatorname{diam} T_i > \varepsilon$.

An immediate consequence is that the number of non-singleton elements in any null collection must be countable. This is easily seen by considering a (countable) sequence $\{\varepsilon_i\}$ converging to zero.



(a) (b)

Figure 9.1. Alternative defining patterns for (a) the Bing decomposition $\mathcal{B}$ and (b) the decomposition $\mathcal{B}_2$.

EXAMPLE 9.2. It is possible to alter the decompositions that we have already encountered, such as the Bing decomposition $\mathcal{B}$ and the decomposition $\mathcal{B}_2$ from Chapter 6, so that they are null. To see this, first recall that a defining sequence for a decomposition can be altered by a homeomorphism at each stage in the sequence (Proposition 4.7), without changing the homeomorphism type of the resulting decomposition space. So we have the freedom to choose the defining sequence asymmetrically, and will obtain a homeomorphic decomposition space. While embedding a sequence of solid tori, we choose one solid torus to be much longer than the other, as in Figure 9.1. More precisely, let T be a standard unknotted solid torus in S^3. Let us construct versions of the $\mathcal{B}$ and $\mathcal{B}_2$ decompositions by embedding the defining patterns shown in Figure 9.1 first into T and then iterating. We have described this process in greater detail earlier, such as in Example 4.6. For convenience, assume that the diameter of T in S^3 is 1. In the ith step of the iterative construction, ensure that the smaller component is embedded with diameter strictly less than 10^{-i}.

Observe that we can organise the defining sequence as a binary tree with one branch for each solid torus: a long branch for each long solid torus and a short branch for each short solid torus. In this scheme, a decomposition element corresponds to an infinite, never backtracking path in the tree. Decomposition elements have nonzero size exactly when we take the short branch only finitely many times. We can estimate the size of those decomposition elements with nonzero diameter: the decomposition elements having diameter at least 10^{-n} for some $n \geq 1$ correspond to those where we take the long branch on the ith iteration for $i \geq n$. This implies that the (repositioned) $\mathcal{B}$ and $\mathcal{B}_2$ decompositions are null. In particular, this shows that there exist null decompositions which shrink, such as the Bing decomposition, and those which do not, such as the $\mathcal{B}_2$ decomposition.

Null decompositions enjoy the following property. Recall from Definition 4.9 that a decomposition $\mathcal{D}$ of a space X is *upper semi-continuous* if each decomposition element Δ is closed and compact, and for each open subset $U \subseteq X$ the set U^* is also open, where U^* is the largest $\mathcal{D}$-saturated subset of U. It is preferable to work with upper semi-continuous decompositions of compact metric spaces since the corresponding decomposition spaces are also compact metric spaces (Corollary 4.13).

LEMMA 9.3. *A null decomposition $\mathcal{D}$ for a metric space (X, d) is upper semi-continuous if each $\Delta \in \mathcal{D}$ is compact.*

PROOF. Since every metric space is Hausdorff, we know that each Δ is closed, since it is compact. By adding singleton decomposition elements, we will assume without loss of generality that the decomposition elements partition the set X.

Let U be an open set in X. Pick $x \in U^*$. We will show that x has an open neighbourhood in U^*, which will imply that U^* is open, completing the proof. The set U^* is $\mathcal{D}$-saturated, so $x \in \Delta \subseteq U^* \subseteq U$, where $\Delta \in \mathcal{D}$ (possibly a singleton). First, we claim there exists an open set $V \supseteq \Delta$ with the property that whenever $\Delta' \cap V \neq \emptyset$, we have $\Delta' \subseteq U$ for any $\Delta' \in \mathcal{D}$. To show this, suppose the converse for a contradiction. That is suppose no such V exists. Define

$$\varepsilon := \min \left\{ \inf_{x \in \Delta, y \in X \smallsetminus U} d(x, y), 1 \right\}.$$

Note that $\varepsilon > 0$ since Δ is compact and $X \smallsetminus U$ is closed. Define open neighbourhoods $N_i := N(\Delta; \varepsilon/2^i) \subseteq U$, for $i \in \{1, 2, \ldots\}$. Then, by assumption, there exists a sequence $\{\Delta_i\}_{i>0}$ of decomposition elements, such that $\Delta_i \cap N_i \neq \emptyset$ and such that $\Delta_i \not\subseteq U$. The choice of N_i implies that $\{\Delta_i\}$ has infinitely many distinct elements since each element of $\mathcal{D}$ is compact. But this in turn implies that the diameters

$\operatorname{diam} \Delta_i > \varepsilon/2$ for all $i > 0$, contradicting the nullity of $\mathcal{D}$. Hence the claimed V exists.

Now we claim that $V \subseteq U^*$. This is immediate from the construction of V and the definition of U^*. That is, since $\mathcal{D}$ is a partition of X, for $v \in V \subseteq X$, there is some $\Delta' \in \mathcal{D}$ with $v \in \Delta'$. Then $\Delta' \cap V \neq \emptyset$ and so $\Delta' \subseteq U^*$. □

For the rest of this chapter $D^n \subseteq \mathbb{R}^n$ will be the closed unit disc around the origin and S^{n-1} will denote its boundary. For a metric space X, recall that a function $f\colon X \to \mathbb{R}$ is said to be *upper semi-continuous* if, for all $x_0 \in X$, we have $\limsup_{x\to x_0} f(x) \leq f(x_0)$, where

$$\limsup_{x\to x_0} f(x) = \lim_{\varepsilon\to 0}(\sup\{f(x) \mid x \in B(x_0;\varepsilon) \smallsetminus \{a\}\}).$$

Informally, this says that f is allowed to jump up but not down. Indicator functions for closed sets are upper semi-continuous.

Definition 9.4. A subset E of $\operatorname{Int} D^n$ is said to be *starlike* if there is a point $O_E \in E$ and an upper semi-continuous function $\rho_E\colon S^{n-1} \to [0,\infty)$ such that $E = \{O_E + t\xi \mid \xi \in S^{n-1},\, 0 \leq t \leq \rho_E(\xi)\}$. Then O_E and ρ_E are called the *origin* and the *radius function* for E respectively.

Example 9.5. There are many equivalent definitions of a starlike set. For example it is clear that a starlike set E in $\operatorname{Int} D^n$ is a compact subset E of $\operatorname{Int} D^n$ which contains a point O_E such that each geometric ray in $\mathbb{R}^n$ emanating from O_E intersects E in a connected set. But conversely, given such a set, take any ray $R(\xi) = \{O_E + t\xi \mid t \in [0,\infty), \xi \in S^{n-1}\}$, and write $\rho_E(\xi) := \sup\{||x - O_E|| \mid x \in R(\xi) \cap E\}$. The compactness of E implies that each ray segment achieves its supremum in E. Thus $E = \{O_E + t\xi \mid \xi \in S^{n-1},\, 0 \leq t \leq \rho_E(\xi)\}$ and $\rho_E\colon S^{n-1} \to [0,\infty)$ is a well defined function. We claim it is moreover upper semi-continuous.

Suppose the function ρ_E is not upper semi-continuous. Then there exists some $\xi_0 \in S^{n-1}$ and some $\varepsilon > 0$ such that every neighbourhood of ξ_0 contains some element $\xi \in S^{n-1}$ with $\rho_E(\xi) > \rho_E(\xi_0) + \varepsilon$. Thus, there is some sequence $\{\xi_i\} \subseteq S^{n-1}$ converging to ξ_0 with $\rho_E(\xi_i) > \rho_E(\xi_0) + \varepsilon$ for all i. Pass to a convergent subsequence of the bounded sequence $\{\rho_E(\xi_i)\}$. Then we assume that there is a sequence $\{\xi_i\}$ converging to ξ_0 with $\{\rho_E(\xi_i)\}$ converging to T for some $T \in \mathbb{R}$ with $T \geq \rho_E(\xi_0) + \varepsilon$. Now note that $\{O_E + \rho_E(\xi_i)\xi_i\} \subseteq E$ is a sequence in the sequentially compact space E. Pass again to a convergent subsequence. Then it is straightforward to see that the subsequence of $\{O_E + \rho_E(\xi_i)\xi_i\}$ converges to $O_E + T\xi_0$. Since E is compact, $O_E + T\xi_0 \in E$ but $T > \rho_E(\xi_0)$, contradicting the definition of ρ_E.

If a starlike set in $\operatorname{Int} D^n$ admits a radius function which is nowhere vanishing and everywhere continuous, then that starlike set is homeomorphic to D^n via the obvious radial dilation from the origin O_E. We call such a starlike set a *starlike ball*. We also have the following useful proposition.

Proposition 9.6. *Every starlike set $E \subseteq \operatorname{Int} D^n$ is cellular.*

Proof. Every bounded upper semi-continuous function ρ is the pointwise limit of a non-increasing sequence of continuous functions $\{\rho^i\}_{i\geq 1}$. For a starlike E in $\operatorname{Int} D^n$, take any origin and radius function (O_E, ρ_E) and let $\rho_E^i\colon S^{n-1} \to [0,\infty)$ for each $i \geq 1$ be such a sequence. It is straightforward to see that these functions may be assumed to be nowhere vanishing for all i. Thus each starlike set $E^i := \{O_E + t\xi \mid \xi \in S^{n-1},\, 0 \leq t \leq \rho_E^i(\xi)\}$ for $i \geq 1$ is a starlike ball. This shows that the set E is cellular and indeed the intersection of starlike balls $E = \bigcap_{i\geq 1} E^i$. □

The shrinking result that we are after will be phrased not in terms of starlike sets, but in terms of the following generalised notions.

DEFINITION 9.7. A subset E of a topological space X is said to be *starlike-equivalent* if, for some n, there is a map $f\colon N(E) \to \operatorname{Int} D^n$ from a closed neighbourhood of E, which is a homeomorphism to its image and such that $f(E)$ is starlike. An *origin* for such an E is $f^{-1}(O)$, where O is an origin for $f(E)$. Note that f, and thus the origin, need not be unique.

A subset E of a topological space X is said to be *recursively starlike-equivalent* (with filtration length K) if there exists a finite filtration

$$\{e \in E\} = E_{K+1} \subseteq E_K \subseteq E_{K-1} \subseteq \cdots \subseteq E_1 \subseteq E_0 = E$$

by closed subsets E_i, such that $E_i/E_{i+1} \subseteq X/E_{i+1}$ is starlike-equivalent for each $0 \leq i \leq K$ and such that E_{i+1}/E_{i+1} may be taken to be the origin of E_i/E_{i+1}.

Recursively starlike-equivalent sets were previously called *eventually starlike-equivalent sets* in [**FQ90**, p. 78].

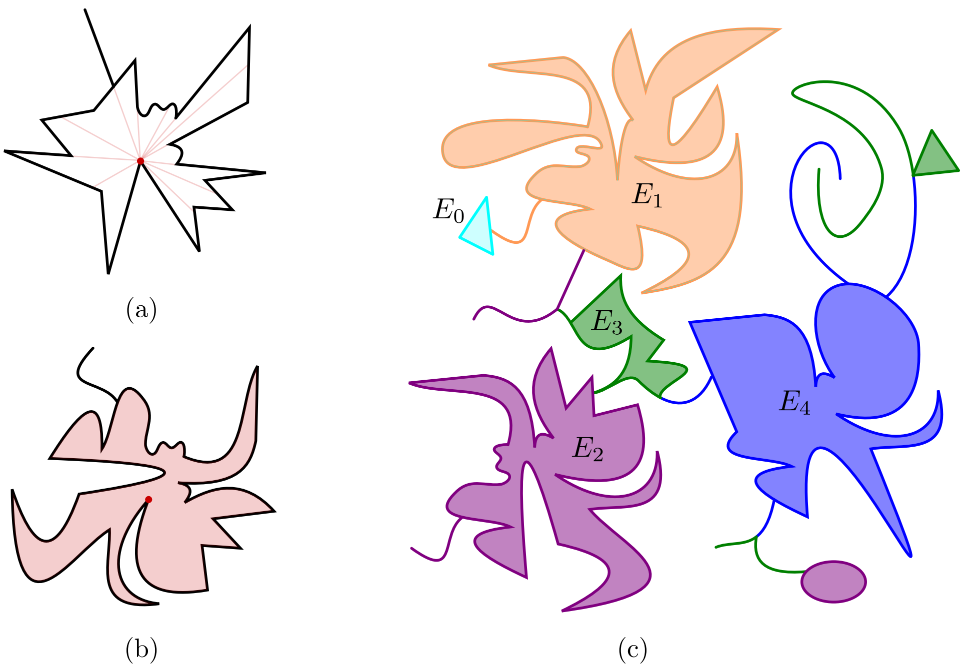


Figure 9.2. Examples of (a) starlike, (b) starlike-equivalent, and (c) recursively starlike-equivalent sets. In (b), the origin is shown in red. In (c), E_4 is shown in dark blue, E_3 is the union of E_4 and the space shown in green, E_2 is the union of E_3 and the space shown in purple, and E_1 is the union of E_2 and the space shown in orange. The entire space is $E_0 = E$.

REMARK 9.8. It is easy to see that a decomposition consisting of a single starlike-equivalent set E in a compact metric space shrinks; in particular, we can use the radial shrink, via the map $f\colon N(E) \to \operatorname{Int} D^n$. Indeed, any finite collection of pairwise disjoint starlike-equivalent sets in a manifold shrinks by Proposition 9.6, Proposition 3.16, and Proposition 3.26.

REMARK 9.9. *A priori* if a set $E \subseteq X$ is recursively starlike-equivalent as in Definition 9.7 then for $i = 0, 1, 2, \ldots, K$ there exist $n_i \in \mathbb{N}$ and maps $f_i \colon N(E_i/E_{i+1}) \to D^{n_i}$ such that f_i is a homeomorphism to its image and that the image of E_i/E_{i+1} is starlike. In fact, these n_i must all be equal as we now show.

Fix $i \in \{1, 2, \ldots, K\}$. The set $E_i/E_{i+1} \subseteq X/E_{i+1}$ is shrinkable since it is starlike-equivalent by definition (see Remark 9.8). Thus the quotient map

$$\pi \colon X/E_{i+1} \to (X/E_{i+1})/(E_i/E_{i+1}) = X/E_i$$

is approximable by homeomorphisms. Note that $\pi(E_{i+1}/E_{i+1}) = E_i/E_i$. Choose $\varepsilon > 0$ small enough so that the ball $B(E_i/E_i; \varepsilon)$ lies within $N(E_{i-1}/E_i)$. Choose a homeomorphism g, approximating π, so that $g(E_{i+1}/E_{i+1}) \in B(E_i/E_i; \varepsilon)$. Let V be a neighbourhood of E_i/E_{i+1} in X/E_{i+1}. Then $g(V)$ intersects the set $N(E_{i-1}/E_i)$ since $g(V)$ contains $g(E_{i+1}/E_{i+1}) \in B(E_i/E_i; \varepsilon) \subseteq N(E_{i-1}/E_i)$.

Thus there is some nonempty open set U contained in the intersection of the sets $g(N(E_i/E_{i+1}))$ and $N(E_{i-1}/E_i)$. Via f_{i-1}, this U is homeomorphic to a nonempty open subset of $D^{n_{i-1}}$ and via $f_i \circ g^{-1}$ it is homeomorphic to a nonempty open subset of D^{n_i}. By invariance of domain, $n_i = n_{i-1}$ for all $i \in \{1, 2, \ldots, K\}$ as claimed.

The following lemma will be used later on in this chapter. Recall that a continuous function $f \colon X \to Y$ between topological spaces is said to be a *quotient map* if it is surjective and a set $U \subseteq Y$ is open if and only if $f^{-1}(U)$ is open.

LEMMA 9.10. *Let E be a recursively starlike-equivalent set with filtration length K in a compact metric space X. Let Y be a topological space. Let $h \colon X \to Y$ be a quotient map which is injective on some open neighbourhood $U \supseteq E$. Then $h(E) \subseteq Y$ is recursively starlike-equivalent with filtration length K.*

PROOF. By hypothesis, there exists a filtration

$$\{e \in E\} = E_{K+1} \subseteq E_K \subseteq E_{K-1} \subseteq \cdots \subseteq E_1 \subseteq E_0 = E$$

where each E_i is closed, and there exist closed neighbourhoods $N(E_i/E_{i+1}) \subseteq X/E_{i+1}$ and functions $f_i \colon N(E_i/E_{i+1}) \to D^n$ for some fixed n which are homeomorphisms onto their image and send E_i/E_{i+1} to a starlike set. By choosing smaller sets if necessary, assume that $N(E_i/E_{i+1}) \subseteq U/E_{i+1}$. This uses the fact that E_i/E_{i+1} is closed in X/E_{i+1} and every metric space is normal [**Mun00**, Theorem 32.2]. Recall that a normal space X satisfies that for every pair A, B of closed sets in X there exist disjoint open neighbourhoods of A and B, which implies that for every closed $A \subseteq X$ and open neighbourhood $A \subseteq V$, there is an open W with $A \subseteq W \subseteq \overline{W} \subseteq V$.

Next, observe that the function h descends to give the quotient maps

$$h_i \colon X/E_{i+1} \to Y/h(E_{i+1}).$$

Then the restrictions $h_i|_{U/E_{i+1}} \colon U/E_{i+1} \to h(U)/h(E_{i+1})$ are bijective quotient maps and thus homeomorphisms. Now we define the filtration

$$\{h(e) \in h(E)\} = h(E)_{K+1} \subseteq h(E)_K \subseteq h(E)_{K-1} \subseteq \cdots \subseteq h(E)_1 \subseteq h(E)_0 = h(E)$$

where we set $h(E)_i = h(E_i)$ for all i. Define the neighbourhood $N(h(E)_i/h(E)_{i+1})$ as the set $h_i(N(E_i/E_{i+1}))$ for all i. Moreover, we have the function

$$f_i \circ h_i|_{U/E_{i+1}}^{-1} \colon N(h(E)_i/h(E)_{i+1}) \to D^n$$

which is a homeomorphism onto its image and maps $h(E)_i/h(E)_{i+1}$ to a starlike set by construction, for each i. This completes the proof. $\square$

In Part IV, we will be interested in recursively starlike-equivalent sets of the following form. Consider an embedded 4-dimensional solid torus $S^1 \times D^3 \subseteq M$ in a 4-manifold M, together with a flat properly embedded disc $D^2 \subseteq \overline{M \smallsetminus (S^1 \times D^3)}$, such that $\partial D^2 = S^1 \times \{1\} \subseteq S^1 \times \partial D^3 \cong S^1 \times S^2$, and such that the flat neighbourhood $D^2 \times D^2$ restricts to a flat neighbourhood in $S^1 \times S^2 = \partial S^1 \times D^3$ for $S^1 \times \{1\}$. Call the union $S^1 \times D^3 \cup D^2$ a *red blood cell* (see Figure 9.3). The flat disc D^2 is called a *red blood cell disc*.

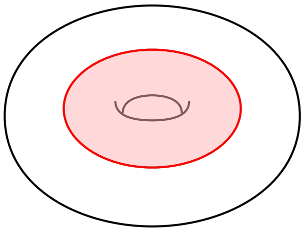


Figure 9.3. A red blood cell with one dimension in the $S^1 \times D^3$ piece suppressed.

Lemma 9.11. *Let $E := S^1 \times D^3 \cup D^2$ be a red blood cell in a 4-manifold M. If $S^1 \times D^3$ has a product neighbourhood $S^1 \times S^2 \times [0,1]$ of its boundary embedded in M, then E is recursively starlike-equivalent.*

Proof. A standard model of a red blood cell consists of a copy of $S^1 \times D^3 \subseteq \mathbb{R}^4$ with core a round circle lying on the xy-plane, with a round disc in the xy-plane glued in. A 4-ball neighbourhood of the standard model is shown on the left of Figure 9.4.

Consider the red blood cell disc $E_1 := D^2 \subseteq \overline{M \smallsetminus (S^1 \times D^3)}$. We have a filtration

$$E_2 = \{\mathrm{pt}\} \subseteq E_1 \subseteq E_0 := E.$$

Write

$$\partial_1 = S^1 \times S^2 \subseteq \partial \overline{M \smallsetminus (S^1 \times D^3)};$$

if M is closed then $\partial_1 = \partial \overline{M \smallsetminus (S^1 \times D^3)}$. By the definition of a red blood cell, the flat neighbourhood of E_1 in an annular neighbourhood of $\partial E_1 \subseteq E_1$ gives a collar of ∂_1 in a neighbourhood of $\partial E_1 \subseteq S^1 \times S^2$. Extend this to a collar neighbourhood of ∂_1, using [**Bro62**, **Arm70**]. By [**Arm70**] (Theorem 3.13), the collar $S^1 \times S^2 \times [0,1]$ provided in the hypothesis is isotopic to this collar.

We summarise the situation. Consider the neighbourhood of the flat disc E_1 as $D^2 \times D^2 \cong E_1 \times D^2$. There is an $\varepsilon > 0$ with $S^1 \times [1-\varepsilon, 1] \times D^2$ extending to a collar neighbourhood of ∂_1 homeomorphic to $S^1 \times S^2 \times [0,1]$, where the two neighbourhoods are identified via

$$\begin{aligned} S^1 \times [1-\varepsilon, 1] \times D^2 &\hookrightarrow S^1 \times S^2 \times [0,1] \\ (x,t,y) &\mapsto (x, y, (t-1+\varepsilon)/\varepsilon). \end{aligned}$$

Here the D^2 in the domain is identified with a subset of S^2 in the codomain such that $S^1 \times D^2$ is a neighbourhood in $S^1 \times S^2$ of ∂E_1.

The union

$$N(E) := S^1 \times D^3 \cup S^1 \times S^2 \times [0,1] \cup E_1 \times D^2 \cong D^4 \subseteq M$$

is then an embedding in M of the 4-ball neighbourhood of the standard model of a red blood cell disc from Figure 9.4. Let $f \colon N(E) \to \mathbb{R}^4$ denote the inverse of this embedding. The red blood cell disc E_1 is sent by f to a round, in particular, starlike disc in a standard coordinate plane. Therefore using $N(E_1) = N(E)$ we see that E_1 is starlike-equivalent.

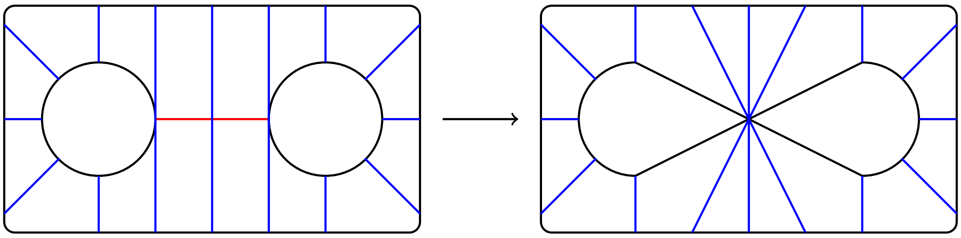


Figure 9.4. Shrinking a red blood cell disc (red). The effect on the standard neighbourhood (blue) of shrinking the red blood cell disc is shown. The red blood cell is shown in cross section.

After we identify E_1 to a point, in the standard model we have that $f(E_0)/f(E_1)$ is starlike in $\mathbb{R}^4/f(E_1) \cong \mathbb{R}^4$, as shown on the right of Figure 9.4. We need to see that $E_0/E_1 = E/E_1$ is starlike-equivalent in M/E_1. For this we use the quotient $N(E_0) := N(E)/E_1$. Since E_1 is starlike-equivalent, it is shrinkable by Remark 9.8. Therefore this neighbourhood is again homeomorphic to D^4 so gives a neighbourhood of E_0/E_1 suitable to show that E_0/E_1 is starlike-equivalent. That is, there is a map $g\colon D^4 = N(E_0) \to \mathbb{R}^4$ sending E_0/E_1 to the starlike standard model shown on the right of Figure 9.4.

Alternatively, once we have the embedding of a 4-ball neighbourhood of the standard model in M, the set E is recursively starlike-equivalent since the standard model is recursively starlike-equivalent by Lemma 9.10, as an embedding is in particular a quotient map onto its image. □

In the standard model of the red blood cell described in the proof above, the red blood cell disc is attached to an unknotted $S^1 \times D^3 \subseteq \mathbb{R}^4$ along a round circle lying in the xy-plane. However, by the light bulb trick, there is a single isotopy class of (unoriented) knots in $S^1 \times S^2$ intersecting a separating sphere $\{*\} \times S^2$ precisely once, and element of this isotopy class may bound a red blood cell disc. Moreover, any two framings of the knot $S^1 \times \{*\} \subseteq S^1 \times S^2$, i.e. identifications of its normal bundle with $S^1 \times D^2$, can be related by a self-homeomorphism of $S^1 \times D^3$. In particular, in Section 28.5 we will find red blood cells within the standard 2-handle $D^2 \times D^2$, where the red blood cell discs are attached along ± 2-framed longitudes of $S^1 \times D^2 \subseteq S^1 \times S^2 = \partial S^1 \times D^3$. These red blood cell discs, together with their 4-ball neighbourhoods, can be identified via a homeomorphism with the standard model red blood cells, so are recursively starlike-equivalent.

9.2. Shrinking null, recursively starlike-equivalent decompositions

We saw earlier that any single starlike-equivalent set in a compact metric space is shrinkable (Remark 9.8). With a little more care, it is also easy to see that any single starlike-equivalent set is also *strongly shrinkable* (Definition 4.16). Now we prove that moreover, given a null collection of sets in the complement of a starlike-equivalent set E, a strong shrinking homeomorphism may be chosen so as to not increase the diameters of the sets in that null collection. Precisely, we begin with the following lemma for starlike sets, the proof of which follows closely that of [**Fre82a**, Lemma 7.1] (see also [**Dav07**, Theorem 8.5]).

Lemma 9.12 (Starlike shrinking lemma). *Let $(E, U, \mathcal{T}, \varepsilon)$ be a 4-tuple consisting of a starlike set $E \subseteq \operatorname{Int} D^n$, a null collection of closed sets $\mathcal{T} = \{T_j\}_{j\geq 1}$ in $\operatorname{Int} D^n \smallsetminus$*

E, an open set $U \subseteq \operatorname{Int} D^n$ containing E, and a real number $\varepsilon > 0$. Then there is a self–homeomorphism h of D^n such that

(1) *h is the identity on $D^n \smallsetminus U$,*
(2) *$\operatorname{diam} h(E) < \varepsilon$, and*
(3) *for each $j \geq 1$, either $\operatorname{diam} h(T_j) < \varepsilon$ or h fixes T_j pointwise.*

Proof. Without loss of generality we assume that the origin O_E of E is the origin of $D^n \subseteq \mathbb{R}^n$. Choose an integer $L > 8/\varepsilon$. For $i = 1, \dots, L$, let B_i be a closed round ball of radius i/L around the origin. Let $V_L \subseteq U$ be a starlike ball containing E and such that $T_j \cap V_L \neq \emptyset$ implies $\operatorname{diam} T_j < \varepsilon/2$. Such a V_L exists since there are only finitely many elements of the null collection with diameter at least $\varepsilon/2$. Since E is the infinite intersection of starlike balls (see proof of Proposition 9.6), we can find starlike balls $\{V_i\}$, for $i = 1, 2, \dots, L-1$, such that $V_i \subseteq \operatorname{Int} V_{i+1}$ and $E \subseteq V_i$ for all i. We can further ensure, by the nullity of $\{T_j\}_{j\geq 1}$, that V_i is small enough to be disjoint from every T_j that intersects the frontier of V_{i+1}. See Figure 9.5.

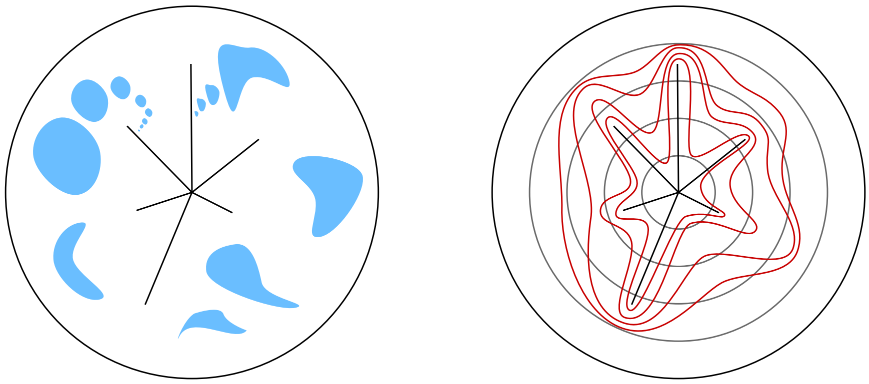


Figure 9.5. Left: A null collection of sets (blue) away from a starlike set E (black) in $\operatorname{Int} D^n$.
Right: Nested starlike neighbourhoods (red) used in the proof of Lemma 9.12. The round balls are shown in grey.

Let $W_i = B_i \cup V_i$. Note that $\{W_i\}_{i=1,\dots,L}$ are nested starlike balls centred at the origin such that $E \subseteq \operatorname{Int} W_i$ for all i. Let $h\colon D^n \to D^n$ be the map determined by the following properties. The map h is defined to be the identity on $D^n \smallsetminus V_L$ and for any ray R in D^n emanating from the origin, the function h is piecewise linear on R and $h(R) = R$. Moreover, for all $i = 1, \dots, L-1$, the function h maps W_i homeomorphically onto B_i. Thus for any point on a ray in $\operatorname{Int} D^n$, h either fixes that point or moves it along that ray closer to the origin. Since the W_i are starlike balls, their radial functions are continuous, which is enough to ensure that this h is a homeomorphism on D^n. Moreover these properties show that h fixes the origin and that $\operatorname{diam} h(E) < \operatorname{diam} B_1 = 1/L < \varepsilon/8 < \varepsilon$. In addition, h is defined to be the identity outside $V_L \subseteq U$. Thus conditions (1) and (2) of the lemma are satisfied. It remains to be shown that this h has the effect required by (3) on the collection $\{T_j\}$.

Lemma 9.13. *Let $i \in \{1, \dots, L\}$ and suppose that $z \in B_i \smallsetminus V_i$. Then $d(z, h(z)) < \varepsilon/8$.*

Proof. For the case $i = L$ this is clear since h fixes such z. Now suppose $i \neq L$, let Rz be the ray from the origin through z, and assume $z \notin B_{i-1}$. (Note

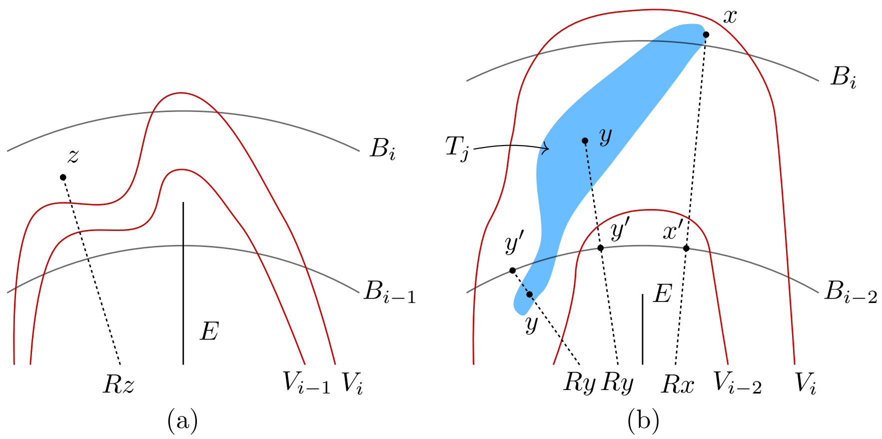


Figure 9.6. (a) How $z \in B_i \smallsetminus V_i$ gets moved by h. (b) How a small T_j gets moved by h. The case of $y \in B_{i-2}$ and $y \notin B_{i-2}$ are both shown.

that $z \in B_i \smallsetminus V_i$, so if $z \in B_{i-1}$, then $z \in B_{i-1} \smallsetminus V_{i-1}$. So assume that i is the smallest such that $z \in B_i \smallsetminus V_i$.) Then the interval $Rz \cap (W_i \smallsetminus W_{i-1})$ is a sub-interval of $Rz \cap (B_i \smallsetminus B_{i-1})$, the latter of which has length less than $\varepsilon/8$ by construction. Since h moves z within this larger interval, z is moved less than $\varepsilon/8$. See Figure 9.6(a). This completes the proof of the lemma. $\square$

We return now to the proof of the proposition. Consider the effect of h on $\{T_j\}$. If $T_j \cap V_L = \emptyset$, then h fixes T_j pointwise by definition. Now suppose $T_j \cap V_L \neq \emptyset$. Then $\operatorname{diam} T_j < \varepsilon/2$ by hypothesis. Suppose that x and y are points in T_j. If $x, y \in B_i \smallsetminus V_i$ for some i, then each is moved less than $\varepsilon/8$ by the lemma above, so the distance between $h(x)$ and $h(y)$ is less than $2(\varepsilon/8) + \varepsilon/2 < \varepsilon$ as needed. Now assume that there is no i such that $x \in B_i \smallsetminus V_i$. By construction, there exists $i \in \{2, \dots, L+1\}$ such that $T_j \subseteq \operatorname{Int} V_i \smallsetminus V_{i-2}$, where we set $V_0 = \emptyset$ and $V_{L+1} = D^n$. In particular, $x, y \in \operatorname{Int} V_i \smallsetminus V_{i-2}$. Note that this forces $x \notin B_{i-2}$ since otherwise $x \in B_{i-2} \smallsetminus V_{i-2}$.

There are now two subcases to consider: either $y \in B_{i-2}$ or $y \notin B_{i-2}$. Let Rx and Ry denote the rays from the origin through x and y, and let x' and y' denote the intersections of these rays with ∂B_{i-2}. See Figure 9.6(b).

If $y \notin B_{i-2}$, then $d(x', y') < d(x, y) < \varepsilon$, since x' and y' are closer to the origin than x and y and lie on a round sphere. Moreover, as $x, y \in V_i \subseteq W_i$, we know that $h(x), h(y) \in B_i$ and $d(h(x), x'), d(h(y), y') < 2(\varepsilon/8) = \varepsilon/4$. It follows that

$$d(h(x), h(y)) \leq d(h(x), x') + d(h(y), y') + d(x', y') < \varepsilon/4 + \varepsilon/4 + \varepsilon/2 = \varepsilon.$$

If $y \in B_{i-2}$, then $d(x, h(y)) \leq d(x, y) + d(y, h(y)) < \varepsilon/2 + \varepsilon/8 = 5(\varepsilon/8)$, since $y \in B_{i-2} \smallsetminus V_{i-2}$ and using our lemma. But $x \notin B_{i-2}$, meaning by the definition of h that $h(x)$ is closer to B_{i-2}, and so to $h(y) \in B_{i-2}$, than x is. Thus, $d(h(x), h(y)) < d(x, h(y)) < 5(\varepsilon/8) < \varepsilon$ as desired. $\square$

Unsurprisingly, with some adjustments to the proof of Lemma 9.12, one can construct a similar careful shrink under the weaker assumption that E is starlike-equivalent, as we show next.

Lemma 9.14. *Let X be a compact metric space and $(E, U, \mathcal{T}, \varepsilon)$ be a 4-tuple consisting of a starlike-equivalent set $E \subseteq X$, an open set $U \subseteq X$ containing E,* *a null collection of closed sets $\mathcal{T} = \{T_j\}_{j\geq 1}$ in $X \smallsetminus E$, and a real number $\varepsilon > 0$. Then there is a self-homeomorphism h of X such that*

(1) *h is the identity on $X \smallsetminus U$,*
(2) *$\operatorname{diam} h(E) < \varepsilon$, and*
(3) *for each $j \geq 1$, either $\operatorname{diam} h(T_j) < \varepsilon$ or h fixes T_j pointwise.*

Proof. By definition, there exists a function $f\colon N(E) \to D^n$ from a closed neighbourhood of E in X which is a homeomorphism to its image and such that $f(E)$ is starlike. Assume, by choosing a slightly smaller $N(E)$ if necessary, that $N(E) \subseteq U$. This again uses the fact that a metric space is normal.

Since X is compact and each element of $\mathcal{T}$ is closed, we see that each element of $\mathcal{T}$ is compact. It is also clear that any starlike-equivalent set is compact (see Example 9.5). Then we see that the decomposition $\mathcal{T} \cup \{E\}$ is null and thus upper semi-continuous by Proposition 9.3. Consequently, there exists some open set V such that $E \subseteq V \subseteq N(E)$ and V is saturated with respect to $\mathcal{T}$. In other words, for any $T \in \mathcal{T}$, $T \cap V \neq \emptyset$ implies that $T \subseteq V$.

Since $N(E)$ is closed in the compact space X, $N(E)$ is compact and thus so is $f(N(E))$. By the Heine-Cantor theorem (Theorem 3.25), the function f as well as the function f^{-1} on $f(N(E))$ is uniformly continuous. In particular, there exists $\delta > 0$ such that for $x, y \in f(N(E))$, if $d(x, y) < \delta$ then $d(f^{-1}(x), f^{-1}(y)) < \varepsilon$.

Let $J \subseteq \mathbb{N}$ be the set such that $j \in J$ implies that $T_j \cap V \neq \emptyset$ (or equivalently, $T_j \subseteq V$). By the uniform continuity of f, the collection $\{f(T_j) \mid j \in J\}$ is a null collection in $\operatorname{Int} D^n \smallsetminus f(E)$.

Write $h'\colon D^n \to D^n$ for the homeomorphism obtained by applying Lemma 9.12 to the 4–tuple $(f(E), f(V), \{f(T_j) \mid j \in J\}, \delta)$, where δ is the value obtained from the Heine-Cantor theorem above. Since h' is the identity on $f(N(E)) \smallsetminus f(V)$, the self-homeomorphism $h := f^{-1} \circ h' \circ f$ of $N(E)$ extends to a self-homeomorphism h of X by declaring h to be the identity on $X \smallsetminus N(E)$. By construction, this h has the required properties (1), (2), and (3). □

Next we prove that null decompositions of a compact metric space consisting of starlike-equivalent sets are shrinkable.

Theorem 9.15. *Suppose that $\mathcal{D}$ is a null decomposition of a compact metric space X consisting of starlike-equivalent sets and that there exists an open set $U \subseteq X$ such that all the non-singleton elements of $\mathcal{D}$ lie in U. Then the quotient map $\pi\colon X \to X/\mathcal{D}$ is approximable by homeomorphisms agreeing with π on $X \smallsetminus U$.*

Proof. Fix $\varepsilon > 0$. Then since $\mathcal{D}$ is null, there exist finitely many decomposition elements $\Delta^1, \Delta^2, \cdots, \Delta^k \in \mathcal{D}$ with $\operatorname{diam} \Delta^i \geq \varepsilon$. We know that $X/\mathcal{D}$ is a compact metric space since $\mathcal{D}$ is upper semi-continuous by Lemma 9.3 (Corollary 4.13). Fix a metric on $X/\mathcal{D}$.

Let $\underline{W}^1$ be an open neighbourhood of $\pi(\Delta^1) \in X/\mathcal{D}$ so that $\operatorname{diam} \underline{W}^1 < \varepsilon$, $\underline{W}^1 \subseteq \pi(U)$, and $\underline{W}^1 \cap \pi(\Delta^j) = \emptyset$ for all $1 < j \leq k$. Let $W^1 = \pi^{-1}(\underline{W}^1)$. Note that W^1 is an open neighbourhood of Δ^1 with $\Delta^1 \subseteq W^1 \subseteq U$ and $W^1 \cap \Delta^j = \emptyset$ for all $1 < j \leq k$. Apply Lemma 9.14 to the 4-tuple $(\Delta^1, W^1, \mathcal{D} \smallsetminus \{\Delta^1\}, \varepsilon)$ to obtain the homeomorphism $h^1\colon X \to X$. Note that h^1 fixes Δ^j pointwise for all $j > 1$, and all decomposition elements in $\mathcal{D} \smallsetminus \{\Delta^i\}_{i=1}^k$ still have diameter less than ε.

Now we iterate. For each i, choose an open neighbourhood $\underline{W}^i$ of $\pi \circ h^{i-1} \circ \cdots \circ h^1(\Delta^i)$ in $X/\mathcal{D}$ so that

(1) $\underline{W}^i \cap \underline{W}^j = \emptyset$ for all $j \leq i - 1$,
(2) $\operatorname{diam} \underline{W}^i < \varepsilon$, and
(3) $\underline{W}^i \subseteq \pi \circ h^{i-1} \circ \cdots \circ h^1(U) = \pi(U)$.

Let $W^i = \pi^{-1}(\underline{W}^i)$. By construction, $W^i \cap W^j = \emptyset$ for all $j \le i-1$ and $\Delta^i \subseteq W^i \subseteq U$. Apply Lemma 9.14 to the 4-tuple $(\Delta^i, W^i, \mathcal{D}^i \smallsetminus \{\Delta^i\}, \varepsilon)$, where $\mathcal{D}^i$ is defined to be $\{h^{i-1} \circ \cdots \circ h^1(\Delta) \mid \Delta \in \mathcal{D}\}$ to produce a homeomorphism $h^i \colon X \to X$. The decomposition $\mathcal{D}^i$ is null since each h^j for $j \le i-1$ is uniformly continuous by the Heine-Cantor theorem (Theorem 3.25). Again, after applying h^i, all decomposition elements in $\mathcal{D}^i \smallsetminus \{\Delta^i\}_{i=1}^k$ still have diameter less than ε by Lemma 9.14 (3).

Now consider the composition $H = h^k \circ \cdots \circ h^1$. By construction, $H|_{X \smallsetminus U}$ is the identity because h^i is the identity on $X \smallsetminus W^i \supseteq X \smallsetminus U$ for each i. We claim that H satisfies the conditions in Definition 4.16. By construction it satisfies condition (ii) that $\operatorname{diam} H(\Delta) < \varepsilon$ for all $\Delta \in \mathcal{D}$. Moreover, for all $x \in X$, either $H(x) = x$ or $x \in W^i$ for some i. If $x \in W^i$ for some i, then $\pi(x), \pi \circ H(x) \in \pi(W^i) = \underline{W}^i$. Then, since $\operatorname{diam} \underline{W}^i < \varepsilon$ by construction, we see that $d(\pi(x), \pi \circ H(x)) < \varepsilon$. Thus, for all $x \in X$, we have that $d(\pi(x), \pi \circ H(x)) < \varepsilon$, or in other words, $d(\pi, \pi \circ H) < \varepsilon$, which is condition (i). As a result, we see that H satisfies both conditions of Definition 4.16, which implies that $\pi \colon X \to X/\mathcal{D}$ is approximable by homeomorphisms restricting π on $X \smallsetminus U$. $\square$

We are now ready to prove our most general shrinking theorem following [**MOR19**, Theorem 1.1].

THEOREM 9.16. *Suppose that $\mathcal{D}$ is a null decomposition of a compact metric space X consisting of recursively starlike-equivalent sets, each of fixed filtration length $K \ge 0$. Suppose that there exists an open set $U \subseteq X$ such that all the non-singleton elements of $\mathcal{D}$ lie in U. Then the quotient map $\pi \colon X \to X/\mathcal{D}$ is approximable by homeomorphisms agreeing with π on $X \smallsetminus U$.*

PROOF. The proof proceeds by induction on K. When $K = 0$, the decomposition elements are all starlike-equivalent, in which case the result follows from Theorem 9.15.

For the inductive step, suppose that the conclusion of the theorem holds when the filtration length is $K-1$ for some $K \ge 1$. Suppose that $\mathcal{D}$ is a null decomposition of X consisting of recursively starlike-equivalent sets of filtration length K. That is, for each $\Delta \in \mathcal{D}$, we have a filtration

$$\{e \in \Delta\} = \Delta_{K+1} \subseteq \Delta_K \subseteq \Delta_{K-1} \subseteq \cdots \subseteq \Delta_1 \subseteq \Delta_0 = \Delta.$$

Consider the decomposition $\mathcal{D}_1 = \{\Delta_1 \mid \Delta \in \mathcal{D}\}$, whose elements are recursively starlike-equivalent of filtration length $K-1$. By the inductive hypothesis, the quotient map $\pi_1 \colon X \to X/\mathcal{D}_1$ is approximable by homeomorphisms restricting to π_1 on $X \smallsetminus U$. Here we are using the fact that $\mathcal{D}_1$ is a null decomposition, since $\mathcal{D}$ is.

Next we consider the decomposition

$$\mathcal{D}/\mathcal{D}_1 := \{\pi_1(\Delta) \mid \Delta \in \mathcal{D}\}$$

of the space $X/\mathcal{D}_1$. Note that the latter is a compact metric space since $\mathcal{D}_1$ is upper semi-continuous by Lemma 9.3 (Corollary 4.13) and moreover, $\pi_1(U)$ is open in $X/\mathcal{D}_1$ (since $\pi_1^{-1}(\pi_1(U)) = U$ is open in X) and contains all the non-singleton elements of $\mathcal{D}/\mathcal{D}_1$. We also note that π_1 is a uniformly continuous function by the Heine-Cantor theorem (Theorem 3.25), and consequently $\mathcal{D}/\mathcal{D}_1$ is a null decomposition, since $\mathcal{D}$ is.

We will next show that $\mathcal{D}/\mathcal{D}_1$ consists of starlike-equivalent subsets of $X/\mathcal{D}_1$. Since each $\Delta \in \mathcal{D}$ is recursively starlike-equivalent, we know by definition that Δ/Δ_1 is starlike-equivalent in X/Δ_1. However, we need to show further that $\pi_1(\Delta)$ is starlike-equivalent in $X/\mathcal{D}_1$.

Fix some $\Delta \in \mathcal{D}$. Observe that $\pi_1 : X \to X/\mathcal{D}_1$ factors as a composition $\pi_1 = \pi_1^b \circ \pi_1^a$ where

$$\pi_1^a : X \to X/\Delta_1$$

and

$$\pi_1^b : X/\Delta_1 \to (X/\Delta_1)/\{\pi_1^a(\Delta_1') \mid \Delta' \in \mathcal{D} \setminus \{\Delta\}\} = X/\mathcal{D}_1$$

are quotient maps. By definition, there exists a closed neighbourhood $N(\pi_1^a(\Delta)) \subseteq X/\Delta_1$ and a function $f : N(\pi_1^a(\Delta)) \to D^n$ for some n which is a homeomorphism onto its image and sends $\pi_1^a(\Delta)$ to a starlike set. Since π_1^a is a uniformly continuous function by the Heine-Cantor theorem (Theorem 3.25), the decomposition

$$\{\pi_1^a(\Delta_1') \mid \Delta' \in \mathcal{D} \setminus \{\Delta\}\}$$

is null in the compact metric space X/Δ_1. By Lemma 9.10, we see that

$$\{\pi_1^a(\Delta_1') \mid \Delta' \in \mathcal{D} \setminus \{\Delta\}\}$$

consists of recursively starlike-equivalent sets of filtration length $K - 1$ and thus by the inductive hypothesis, the function π_1^b is approximable by homeomorphisms which agree with π_1^b on the closed set $\pi_1^a(\Delta)$. Let g be any such approximating homeomorphism. Note that $g(N(\pi_1^a(\Delta))$ is a closed neighbourhood of $g(\pi_1^a(\Delta)) = \pi_1^b(\pi_1^a(\Delta)) = \pi_1(\Delta)$. We also have the function $f \circ g^{-1} : g(N(\pi_1^a(\Delta)) \to D^n$, which is a homeomorphism onto its image and maps $\pi_1(\Delta)$ to a starlike set by construction. This completes the argument that for each $\Delta \in \mathcal{D}$, $\pi_1(\Delta)$ is starlike-equivalent in $X/\mathcal{D}_1$.

We have now shown that the decomposition $\mathcal{D}/\mathcal{D}_1$ and the set $\pi_1(U)$ satisfy the hypotheses of Theorem 9.15 in the space $X/\mathcal{D}_1$. As a result, the quotient map $\pi_0 : X/\mathcal{D}_1 \to (X/\mathcal{D}_1)/(\mathcal{D}/\mathcal{D}_1) = X/\mathcal{D}$ is approximable by homeomorphisms restricting to π_0 on $X/\mathcal{D}_1 \setminus \pi_1(U)$. Then since the composition of maps which are approximable by homeomorphisms is itself approximable by homeomorphisms (Proposition 3.24), we infer that $\pi : X \to X/\mathcal{D}$ is approximable by homeomorphisms restricting to $\pi = \pi_0 \circ \pi_1$ on $X \setminus U$ as desired. By induction, this completes the proof of the theorem. $\square$

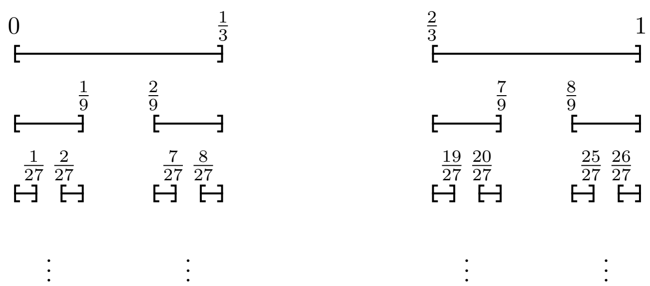


Figure 9.7. Constructing the ternary Cantor set.

Example 9.17. Recall that in Remark 4.15 we considered the decomposition of $[0, 1]$ consisting of the closures of the connected components of the complement of the ternary Cantor set $\mathfrak{C}_3 \subseteq [0, 1]$. All the decomposition elements are intervals in $[0, 1]$ and so are starlike. The decomposition is also null since the intervals removed in the ith step of the construction of the Cantor set have length 3^{-i}. Therefore Theorem 9.16 (or Theorem 9.15) tells us that this is a shrinkable decomposition and so $[0, 1]/([0, 1] \setminus \mathfrak{C}_3) \cong [0, 1]$ as previously claimed.

The result needed for the proof of the disc embedding theorem in Part IV is the following, which is an immediate corollary of Theorem 9.16.

THEOREM 9.18 (Starlike null theorem). *Let $\mathcal{D}$ be a null decomposition of $D^2 \times D^2$ with recursively starlike-equivalent decomposition elements, each of fixed filtration length $K \geq 0$. Suppose the decomposition elements are disjoint from the boundary of $D^2 \times D^2$. Also suppose that $S^1 \times D^2 \subseteq \partial(D^2 \times D^2)$ has a closed neighbourhood C disjoint from the decomposition elements. Then the quotient map $\pi \colon D^2 \times D^2 \to (D^2 \times D^2)/\mathcal{D}$ is approximable by homeomorphisms, each of which agrees with π on $C \cup \partial(D^2 \times D^2)$. In particular, there is a homeomorphism of pairs*

$$h \colon (D^2 \times D^2, S^1 \times D^2) \xrightarrow{\cong} ((D^2 \times D^2)/\mathcal{D}, S^1 \times D^2)$$

restricting to the quotient map π on $C \cup \partial(D^2 \times D^2)$.

9.3. Literature review

Before finishing this chapter we give a quick roundup of the history of the results we have presented. Bean proved in [**Bea67**] that null, starlike decompositions of $\mathbb{R}^3$ shrink and in [**Dav07**, Theorem 8.6], Daverman outlines an extension of this theorem to $\mathbb{R}^n$ for all n. The notion of recursively starlike-equivalent sets was introduced by Denman and Starbird in [**DS83**], where such sets were called *birdlike-equivalent sets*. Our definition is more general than theirs, and was introduced by Freedman and Quinn in [**FQ90**, p. 78], where they were called *eventually starlike-equivalent*. Bing showed that countable, starlike, upper semi-continuous decompositions of $\mathbb{R}^n$ are strongly shrinkable [**Dav07**, Theorem 8.7]. We saw earlier that null decompositions are countable and that a null decomposition consisting of closed and compact elements is upper semi-continuous (Lemma 9.3). Thus a starlike version of Theorem 9.16 for $\mathbb{R}^n$ is obtained as a straightforward consequence of Bing's theorem. A difficulty we have not needed to consider, and that Bing overcame, is that when the decomposition is not null some decomposition elements may get stretched by the shrinking of other elements. Over the course of countably many shrinks it is possible, but becomes somewhat technical, to prevent this phenomenon from concatenating.

A weaker version of Theorem 9.16, namely that the quotient space is homeomorphic to the original space, was left as an exercise in [**FQ90**, Section 4.5]. A stronger version of Theorem 9.16, removing the requirement of a uniform upper bound on the filtration lengths, has been proved by Ancel in [**Anc20**].

In terms of proving the disc embedding theorem in [**Fre82a**], Freedman needed a result that said the "holes$^+$" decomposition of $D^2 \times D^2$ shrinks. The holes$^+$ decomposition in Part IV will consist of the red blood cells of Lemma 9.11. We already saw that red blood cells are recursively starlike-equivalent. When Freedman originally shrunk this decomposition, he did not use a shrinking result as general as Theorem 9.16. He rather argued as far as the statement that null, starlike-equivalent decompositions shrink and then gave an *ad hoc* argument for a 2-stage shrink; see [**Fre82a**, Section 8]. Another *ad hoc* approach is given in [**FQ90**, Section 4.5]. Our choice to present a slightly more general shrinking result was made to emphasise that this particular part of the proof of the disc embedding theorem belongs to a collection of standard ideas from the decomposition space theory literature.

CHAPTER 10

The ball to ball theorem

Stefan Behrens, Boldizsár Kalmár, and Daniele Zuddas

We discuss the final key ingredient from decomposition space theory needed for the proof of the disc embedding theorem, namely the *ball to ball theorem*, which gives a sufficient condition for a map $D^4 \to D^4$ to be approximable by homeomorphisms. We give the formulation from [**FQ90**, p. 80] but a proof which is similar to the original one given in [**Fre82a**]. There are other accounts by Ancel [**Anc84**] and Siebenmann [**Sie82**]. We have previously seen that not all null decompositions shrink. For example, the decomposition $\mathcal{B}_2$ from Chapter 6 is not shrinkable; we explained why it is null in Example 9.2. In Chapter 9, we saw that null decompositions with *recursively starlike-equivalent* elements do shrink. The ball to ball theorem provides another situation wherein a null decomposition shrinks. Given a null decomposition of the ball D^4 such that the decomposition space is homeomorphic to D^4 as well, and the image of the decomposition elements under the quotient map is *nowhere dense*, the theorem states that the quotient map is approximable by homeomorphisms. This is remarkable in that, beyond nullity, we do not need to know anything about the decomposition elements. As mentioned before, we will use this theorem in the very last step of the proof of the disc embedding theorem, namely in the β *shrink*.

Recall that a collection of subsets of a metric space is said to be *null* if for every $\varepsilon > 0$ there are only finitely many elements of diameter greater than ε (Definition 9.1). A subset A of a topological space X is said to be *nowhere dense* if its closure has empty interior. Given a function $f\colon X \to Y$, a set $f^{-1}(y)$ is said to be an *inverse set* of f if $|f^{-1}(y)| > 1$ (Definition 3.21). The image of the inverse sets of f is called the *singular image* of f and is henceforth denoted by $\mathrm{Sing}(f)$.

THEOREM 10.1 (Ball to ball theorem). *Let $f\colon D^4 \to D^4$ be a continuous map restricting to a homeomorphism $f|_{\partial D^4}\colon \partial D^4 \to \partial D^4$, and let E be a closed subset of D^4 containing ∂D^4. Suppose that the following holds.*

(a) The collection of inverse sets of f is null.
(b) The singular image of f is nowhere dense.
(c) The map f restricts to a homeomorphism $f|_{f^{-1}(E)}\colon f^{-1}(E) \to E$.

Then f can be approximated by homeomorphisms that agree with f on $f^{-1}(E)$.

The fact that f restricts to a homeomorphism from $S^3 = \partial D^4$ to itself implies that f has degree one and is thus surjective.

Observe that the inverse sets of f form an upper semi-continuous decomposition $\mathcal{D}$ of D^4 by Lemma 9.3, since every inverse set is closed in the compact space D^4. Let $\pi\colon D^4 \to D^4/\mathcal{D}$ be the quotient map. We obtain a well defined continuous function $\overline{f}\colon D^4/\mathcal{D} \to D^4$ since f is constant on the fibres of π. The function $\overline{f}$ is a continuous bijection from a compact space to a Hausdorff space and is thus a homeomorphism by the closed map lemma (Lemma 3.23). The ball to ball theorem establishes that f is approximable by homeomorphisms $f_n\colon D^4 \to D^4$. By composing with $\overline{f}^{-1}$, we see that the quotient map π is approximable by the homeomorphisms $\overline{f}^{-1} \circ f_n \colon D^4 \to D^4/\mathcal{D}$. This used that $\overline{f}^{-1}$ is uniformly continuous by the Heine-Cantor theorem (Theorem 3.25). The following diagram shows the maps. The diagram commutes when the top arrow denotes the function f.

$$\begin{tikzcd} D^4 \arrow[r, "f", "f_n"'] \arrow[d, "\pi"'] & D^4 \\ D^4/\mathcal{D} \arrow[ur, "\overline{f}"'] & \end{tikzcd}$$

We also remark that the restriction to dimension four is completely irrelevant and the proof we give will work in all dimensions. However, a result of Siebenmann [**Sie73**] shows that functions between manifolds of dimension five or higher, with far fewer restrictions than in Theorem 10.1, are approximable by homeomorphisms. This was known prior to the ball to ball theorem, which would have been an easy special case in high dimensions. Thus the 4-dimensional case is what requires a special argument.

10.1. The main idea of the proof

Before going into the details we outline the main idea of the proof. First, in an attempt to reduce confusion, we relabel the source and the target copies of D^4 as X and Y and consider the function $f \colon X \to Y$.

The proof will directly use the definition of a function being approximable by homeomorphisms (Definition 3.19). That is, given $\varepsilon > 0$, we will find a homeomorphism $h \colon X \to Y$ such that $d(h, f) < \varepsilon$. Note that f fails to be a homeomorphism exactly at the inverse sets and at first sight the ones with largest diameter are the most offensive. With the goal of constructing a homeomorphism h as above, we wish to control the inverse sets with diameter greater than some fixed ε, and by the hypothesis that the collection of inverse sets is null, we have only finitely many such inverse sets to worry about.

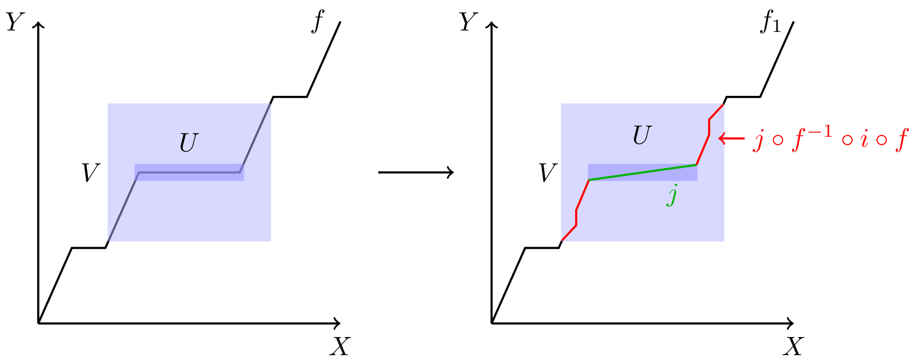


Figure 10.1. A modification of the Cantor function. On the right, we see that certain vertical segments (red) are created.

Consider the Cantor function (Figure 10.1) which has a large horizontal spot in the middle, demonstrating its non-injectivity. We present a strategy to approximate the Cantor function by homeomorphisms. Finding such an approximation for the Cantor function is in general not hard, but the complicated strategy we are about to give has the advantage of working in the 4-dimensional context. We work with the graph of the Cantor function as in Figure 10.1 and find small neighbourhoods $U \subseteq V$ of the region in the graph corresponding to the largest inverse set, namely the horizontal spot in the middle. We then modify the function. It remains the same outside of V. Within U, we tilt it slightly, in order to eliminate the inverse set. However, the price we pay is the creation of certain vertical cliffs in the region $V \setminus U$. In other words, the result of the modification no longer has the large inverse set but is also no longer a function. It is instead a *relation*, meaning that it is a multi-valued function. We will define relations carefully soon. On the other hand, the vertical regions are located in small neighbourhoods and we may hope to be able to control them.

The strategy of the proof is then as follows. We perform the modification above to all the large inverse sets we wish to control. This replaces large horizontal steps by smaller vertical steps. Then we utilize the symmetry of the situation and consider the graph for the inverse function (really, relation). This has a number of vertical steps, upon which we can apply the same modification as above, before using the symmetry again. This process creates some new horizontal steps, but these have small diameter. By performing this iteration over and over again, we eventually reduce all the horizontal regions of our function to small diameter, as desired.

In preparation for the proof, we give the precise formulae for the first step. Assume that some inverse set of f maps to $0 \in Y$, and suppose this is the inverse set we presently wish to control. Take two concentric, closed, round balls $U \subsetneq V \subseteq Y$ with distinct radii around $0 \in Y$, where U has very small radius. Define a stretching homeomorphism $i\colon V \to Y$ that is the identity on U and sends $V \setminus U$ onto $Y \setminus U$. In particular, $i|_{\partial V}\colon \partial V \to \partial Y$ is a homeomorphism. We could define such a function i as a radial expansion for now, but in later steps we will need more flexibility.

Note that the composition $i^{-1} \circ f\colon X \to V$ restricted to ∂X is a homeomorphism onto ∂V. Therefore $(i^{-1} \circ f)|_{\partial X}$ can be extended to a homeomorphism $j\colon X \to V$ by the Alexander trick.

Finally, we define a relation $f_1\colon X \to Y$ as follows.

$$f_1 = \begin{cases} f & \text{on } X \setminus f^{-1}(V) \\ j \circ (f^{-1} \circ i \circ f) & \text{on } f^{-1}(V \setminus U) \\ j & \text{on } f^{-1}(U), \end{cases} \tag{10.1}$$

using that $f(f^{-1}(V \setminus U)) \subseteq V$. We check that the definitions match on the overlaps $\partial f^{-1}(V)$ and $\partial f^{-1}(U)$. Recall that since f is continuous, $\partial f^{-1}(V) \subseteq f^{-1}(\partial V)$ and $\partial f^{-1}(U) \subseteq f^{-1}(\partial U)$. First, $i = \mathrm{Id}$ on U so

$$j \circ f^{-1} \circ i \circ f = j \circ f^{-1} \circ f = j$$

on $f^{-1}(\partial U)$. Next, $j = i^{-1} \circ f$ on ∂X. Therefore

$$j \circ f^{-1} \circ i \circ f = i^{-1} \circ f \circ f^{-1} \circ i \circ f = f$$

on $f^{-1}(\partial V)$.

Observe that f_1 no longer has a singular point at 0, since $f_1^{-1}(0) = j^{-1}(0)$ and j is a homeomorphism.

However f_1 is not a function, as we now explain. Consider a point $y \in V \setminus U \subseteq Y$ such that $i(y)$ lies in the singular image of f. Recall that f is surjective. Then given $x \in f^{-1}(V \setminus U)$ with $y = f(x)$, the set $f^{-1}(i(f(x)))$ has cardinality strictly greater than one. This cardinality is preserved under the homeomorphism j. In other words, the relation f_1 maps x to the set $\{j \circ f^{-1} \circ i \circ f(x)\}$, which has more than one element and as a result f_1 is not a function. Helpfully this failure only occurs for points mapping to V, and we could choose V to be arbitrarily small, as long as it is bigger than U. Thus the places where f_1 fails to be a function can be controlled.

This describes the principal idea behind the proof of the ball to ball theorem. In the next two sections we address some necessary preliminaries.

10.2. Relations

Given sets X and Y, a subset of $X \times Y$ is said to be a *relation* from X to Y. For the rest of this section, we restrict ourselves to relations where for every $x \in X$ there is some $y \in Y$ with $(x, y) \in R$. These are denoted by $R\colon X \to Y$, expressing the fact that the *domain* of R is all of X. We think of a relation $R\colon X \to Y$ as a multi-valued function which assigns to every $x \in X$ the subset $R\{x\} \coloneqq \{y \in Y \mid (x, y) \in R\}$. Of course, any function from X to Y is a relation from X to Y. More precisely, we define the *graph* of a function $f\colon X \to Y$ as the set $\{(x, y) \mid x \in X, y \in Y, y = f(x)\} \subseteq X \times Y$. A function corresponds uniquely to its graph and thus corresponds to a relation. We will frequently refer to the graph of a function as simply the function for convenience. A relation $R\colon X \to Y$ is (the graph of) a function if and only if $R\{x\}$ is a singleton set for all $x \in X$, in which case we write $R(x)$ for the image of $x \in X$, that is, $R\{x\} = \{R(x)\}$ for all $x \in X$.

Several notions for functions are meaningful for relations. We say that a relation $R\colon X \to Y$ is *surjective* if for every $y \in Y$ there is some $x \in X$ such that $(x, y) \in R$. Given sets X, Y, and Z, we define the *composition* of relations $R\colon X \to Y$ and $S\colon Y \to Z$ as

$$S \circ R = \{(x, z) \in X \times Z \mid \text{ there exists } y \in Y \text{ such that } (x, y) \in R \text{ and } (y, z) \in S\}.$$

Similarly, we define the *inverse* of a relation $R\colon X \to Y$ as

$$R^{-1} = \{(y, x) \in Y \times X \mid (x, y) \in R\}$$

and we can write $R^{-1}\colon Y \to X$ if R is surjective. It is straightforward to verify that $(S \circ R)^{-1} = R^{-1} \circ S^{-1}$ and that the notions above correspond to the usual notions for functions.

The following well known lemma (see e.g. [**Mun00**, p. 171]) shows how to detect the continuity of a function by examining its graph.

Lemma 10.2 (Closed graph lemma). *Let X and Y be compact, Hausdorff topological spaces. Then $f\colon X \to Y$ is continuous if and only if the graph of f is a closed subset of $X \times Y$.*

Proof. This is an easy exercise in topology. Let G denote the graph of f. Suppose that f is continuous and that (x, y) is not in G. Then $f(x) \neq y$, and thus $f(x)$ and y can be separated by open sets $U \ni f(x)$ and $V \ni y$ in Y with $U \cap V = \emptyset$. Since f is continuous we can choose an open neighbourhood $W \ni x$ so that $f(W) \subseteq U$. Then $W \times V$ is disjoint from G. To see this, note that if $(x', y') \in W \times V$, then $f(x') \in U$ while $y' \in V$. Since U and V are disjoint, $f(x') \neq y'$. Thus, $W \times V$ is an open neighbourhood for (x, y) away from G. It follows that G is closed.

Assume G is closed in $X \times Y$. Let $C \subseteq Y$ be closed. We will show that $f^{-1}(C)$ is closed in X. Since $X \times Y$ is compact and X is Hausdorff we know that the continuous projection $\pi\colon X \times Y \to X$ is a closed map by the closed map lemma (Lemma 3.23). Then $f^{-1}(C) = \pi((X \times C) \cap G)$ is closed. $\square$

We call a relation *closed* if it is a closed subset of $X \times Y$. By Lemma 10.2, closed relations form an appropriate generalisation of continuous functions.

For metric spaces X and Y, the failure of a relation $R\colon X \to Y$ to be a function is measured by the quantity

$$\operatorname{vd}(R) \coloneqq \sup_{x \in X} \operatorname{diam}_Y R\{x\}$$

which we call the *vertical defect* of R. We also define the *horizontal defect* of a surjective relation $R\colon X \to Y$ by

$$\mathrm{hd}(R) := \sup_{y\in Y} \mathrm{diam}_X\, R^{-1}\{y\} = \mathrm{vd}(R^{-1})$$

which is the obstruction for R^{-1} to be a function. The terminology corresponds to the case of relations $[0,1] \to [0,1]$ as is readily seen by a graph in the traditional sense (look at Figure 10.1 again).

The following proposition establishes precisely when a relation corresponds to a homeomorphism.

Proposition 10.3. *For compact metric spaces X and Y, a relation $R\colon X \to Y$ is a homeomorphism if and only if the following conditions hold.*

(i) *R is a closed relation.*
(ii) *R is a surjective relation.*
(iii) $\mathrm{hd}(R) = \mathrm{vd}(R) = 0$.

Proof. The only if direction is immediate. Since $\mathrm{hd}(R) = \mathrm{vd}(R) = 0$ and R is surjective, both R and R^{-1} are bijective functions. Since R, and thus R^{-1}, is closed, both R and R^{-1} are continuous by the closed graph lemma (Lemma 10.2). This proves the if direction. $\square$

As a final piece of notation we define the *singular image* of a relation $R \subseteq X \times Y$ by

$$\mathrm{Sing}(R) = \{y \in Y \mid |R^{-1}\{y\}| > 1\}.$$

This matches with our previous definition of the singular image of a function.

We finish this section with another word on the proof of the ball to ball theorem. In previous chapters our strategy for shrinking has stayed within the world of continuous functions and we have succeeded by reducing the horizontal defect, although we did not describe it in those exact terms. In the upcoming proof we will stray into the world of relations and try to reduce both the horizontal and vertical defects. Proposition 10.3 and the closed graph lemma (Lemma 10.2) will allow us to return to the world of continuous functions and homeomorphisms once we have succeeded in killing the horizontal and vertical defects.

10.3. Admissible diagrams and the main lemma

In this section we define a tool that will help us keep track of singular images when we iterate the proof. We prove a key technical lemma, Lemma 10.6.

Recall that the idea of the proof is to iterate the construction of f_1 in (10.1). In order to do this, we will extend the construction from functions to relations and keep track of the singular images of both the relation and its inverse. A convenient way of doing this is to use a third copy of D^4 to keep track of where the singular sets lie. The following notion will be helpful.

Definition 10.4 (Admissible diagram). Let X, Y, and Z be three copies of D^4. Let $F \subseteq X$ be a closed neighbourhood of ∂X, let $E \subseteq Y$ be a closed neighbourhood of ∂Y, and let $f\colon F \to E$ be a homeomorphism. A diagram

$$\begin{tikzcd} X \arrow[rr, bend left, "R"] \arrow[dr, "r"'] & & Y \arrow[ll, bend left, "S"] \arrow[dl, "s"] \\ & Z & \end{tikzcd}$$

is said to be *admissible* if:

(a) $R\colon X \to Y$ and $S\colon Y \to X$ are closed, surjective relations;

(b) r and s are continuous functions;
(c) $s \circ R = r$ and $r \circ S = s$, that is the diagram commutes;
(d) $S = R^{-1}$;
(e) the collections of inverse sets of r and s are null (and hence countable, as discussed in Section 9.1);
(f) the singular images $\operatorname{Sing}(r)$ and $\operatorname{Sing}(s)$ are nowhere dense in Z;
(g) R and r are homeomorphisms on F, with $R|_F = f$;
(h) S and s are homeomorphisms on E, with $S|_E = f^{-1}$;
(i) the sets $A := \operatorname{Sing}(r) \smallsetminus \operatorname{Sing}(s)$, $B := \operatorname{Sing}(s) \smallsetminus \operatorname{Sing}(r)$, and $C := \operatorname{Sing}(r) \cap \operatorname{Sing}(s)$ of Z are *mutually separated* in Z, that is each is disjoint from the closures of the others; and
(j) the relation R restricts to a homeomorphism $R|_{r^{-1}(C)} \colon r^{-1}(C) \to s^{-1}(C)$ with inverse $S|_{s^{-1}(C)} \colon s^{-1}(C) \to r^{-1}(C)$.

An admissible diagram as above is denoted by $\mathcal{A} = (R, S; r, s)$.

Observe that a diagram $(R, S; r, s)$ is admissible with respect to f if and only if the diagram $(S, R; s, r)$ is admissible with respect to f^{-1}. Additionally, an admissible diagram $(R, S; r, s)$ is determined by the pair (R, r). We have the following straightforward lemma.

Lemma 10.5. *Let $(R, S; r, s)$ be an admissible diagram. Using the notation in Definition 10.4(i), we have that $s(\operatorname{Sing}(R)) = A$ and $r(\operatorname{Sing}(S)) = B$.*

Proof. We show that $s(\operatorname{Sing}(R)) = A := \operatorname{Sing}(r) \smallsetminus \operatorname{Sing}(s)$. Let $y \in \operatorname{Sing}(R)$, that is, there exist $x_1 \neq x_2 \in X$ such that $y \in R\{x_1\}$ and $y \in R\{x_2\}$. Then $s(y) = r(x_1) = r(x_2)$ since $s \circ R = r$ and thus $s(y) \in \operatorname{Sing}(r)$. Suppose that $s(y) \in \operatorname{Sing}(s)$. Then $s(y) \in \operatorname{Sing}(r) \cap \operatorname{Sing}(s) =: C$. Then $x_1, x_2 \in r^{-1}(C)$ and $y \in s^{-1}(C)$ with $y \in R\{x_1\}$ and $y \in R\{x_2\}$. But this is a contradiction since R is a homeomorphism (and in particular a bijection) on $r^{-1}(C)$ by hypothesis. Therefore $s(\operatorname{Sing}(R)) \subseteq A$.

Now we prove that $A \subseteq s(\operatorname{Sing}(R))$. Let $z \in A := \operatorname{Sing}(r) \smallsetminus \operatorname{Sing}(s)$. Since $z \in \operatorname{Sing}(r)$, there exist $x_1 \neq x_2 \in X$ with $r(x_1) = r(x_2) = z$. Since $s \circ R = r$, we have that $(s \circ R)(x_1) = (s \circ R)(x_2) = z$. Now note that if $R\{x_1\} \neq R\{x_2\}$, then $z \in \operatorname{Sing}(s)$, which is a contradiction. Thus, $R\{x_1\} = R\{x_2\}$. We have that $z = r(x_1) = r(x_2)$. Since $s \circ R = r$, we know that there exists $y \in R\{x_1\} = R\{x_2\}$ such that $s(y) = z$. Since $x_1 \neq x_2$, we see that $z \in s(\operatorname{Sing}(R))$, as needed.

The proof that $r(\operatorname{Sing}(S)) = B$ is directly analogous. $\square$

The starting data for our iterative scheme in the proof of Theorem 10.1 will be the diagram

$$\begin{tikzcd} X \arrow[rr, bend left, "f"] \arrow[dr, "f"'] & & Y \arrow[ll, bend left, "f^{-1}"] \arrow[dl, "\mathrm{Id}"] \\ & Z & \end{tikzcd}$$

with $E \subseteq Y$ the E from the statement of Theorem 10.1, $F = f^{-1}(E) \subseteq X$, and $f \colon F \to E$ the restriction of $f \colon X \to Y$. This is admissible by the assumptions of Theorem 10.1. In particular, we use the fact that f^{-1} is a closed relation since f is. Additionally, since s is the identity map, the sets B and C are empty. The set $A = \operatorname{Sing}(f)$ can be nonempty. In each step of the iteration, we will erase some of the singular image of f but create singular image of the inverse relation, which means that B and C come alive. Thus, we will need to know what to do when all three of A, B, and C are nonempty. Roughly speaking, we will gradually move the points of A into B and C, while making the size of all three smaller.

Now we give the main technical lemma from which the proof of the ball to ball theorem will follow. Recall that a neighbourhood of a subset $A \subseteq X$ in a space is a set $B \subseteq X$ such that there is an open set U with $A \subseteq U \subseteq B$.

LEMMA 10.6. *Let $F \subseteq X$ be a closed neighbourhood of ∂X, let $E \subseteq Y$ be a closed neighbourhood of ∂Y, and let $f\colon F \to E$ be a homeomorphism. Let $(R, S; r, s)$ be an admissible diagram.*

Then for every neighbourhood $\mathcal{N}(R) \subseteq X \times Y$ of R and for every $\varepsilon > 0$ there is an admissible diagram $(R', S'; r', s')$ with $r' = r$, $R'|_F = R|_F$, $R' \subseteq \operatorname{Int} \mathcal{N}(R)$, and $\operatorname{hd}(R') < \varepsilon$.

In other words, there is a relation R' that is arbitrarily close to R, unchanged on F, and whose horizontal spots, namely the inverse sets, have arbitrarily small size. In the proof, we will need the following lemma, whose proof we postpone.

LEMMA 10.12 (General position lemma [**Anc84**, Lemma 1(2)]). *Let Z be a compact manifold with a metric d inducing its topology, and let G be a closed subset of Z. Let $P, Q \subseteq Z \smallsetminus G$ be countable and nowhere dense subsets in the complement of G. For every $\eta > 0$ there exists a self-homeomorphism h_η of Z such that $d(h_\eta, \mathrm{Id}_Z) < \eta$, $h_\eta = \mathrm{Id}$ on G, and such that $h_\eta(P)$ and Q are mutually separated.*

PROOF OF LEMMA 10.6. We will use the notation of Definition 10.4. Let A^ε denote the subset of $A := \operatorname{Sing}(r) \smallsetminus \operatorname{Sing}(s) = s(\operatorname{Sing}(R)) \subseteq Z$ consisting of the points $a \in A$ such that the diameter of $r^{-1}(a)$ is at least ε. Then A^ε is a finite set since the collection of inverse sets of r is null by Definition 10.4(e). We follow the strategy sketched in Section 10.1. Let $U, V \subseteq Z$ be closed balls centred at some $a \in A^\varepsilon$ such that the following is satisfied.

(i) The set $r^{-1}(a)$ has maximal diameter among the inverse sets of r.
(ii) We have that $a \in U \subseteq \operatorname{Int} V$ and $V \subseteq Z \smallsetminus (\overline{\operatorname{Sing}(s)} \cup s(E))$. This is possible to arrange because A is separated from $\overline{\operatorname{Sing}(s)} = \overline{B \cup C}$ by hypothesis and $a \notin s(E)$.
(iii) We have that $\partial U \cap \operatorname{Sing}(r) = \emptyset$. This is possible because $\operatorname{Sing}(r)$ is countable.

See Figure 10.2 for a schematic picture.

By construction, the function s is nonsingular on $s^{-1}(V)$. Then the function $s|_{s^{-1}(V)}\colon s^{-1}(V) \to V$ is a continuous bijection from a compact space to a Hausdorff space and so is a homeomorphism by the closed map lemma (Lemma 3.23). Let

$$i\colon V \to Z$$

be a homeomorphism onto Z such that $i|_U$ is the identity map. Perturb the map i using the general position lemma (Lemma 10.12) in Z, i.e. replace i with $h_\eta \circ i$, taking

$$P := i((\operatorname{Sing}(r) \smallsetminus U) \cap V),\ Q := \operatorname{Sing}(r) \smallsetminus U, \text{ and } G := s(E) \cup U = r(F) \cup U$$

in the notation of Lemma 10.12, to arrange that

$$\begin{aligned} h_\eta \circ i((\operatorname{Sing}(r) \smallsetminus U) \cap V) \cap \overline{(\operatorname{Sing}(r) \smallsetminus U)} &= \emptyset \text{ and} \\ \overline{h_\eta \circ i((\operatorname{Sing}(r) \smallsetminus U) \cap V)} \cap (\operatorname{Sing}(r) \smallsetminus U) &= \emptyset, \end{aligned} \tag{10.2}$$

with $h_\eta \circ i|_U$ still the identity map. In other words, P and Q are mutually separated.

We do not need to control the value of η in the application of the general position lemma. From now on we refer to the perturbed map $h_\eta \circ i$ again as i, absorbing the perturbation homeomorphism into the notation. This map is the analogue of the radial stretching map, also called i, from Section 10.1. We record the following observation.

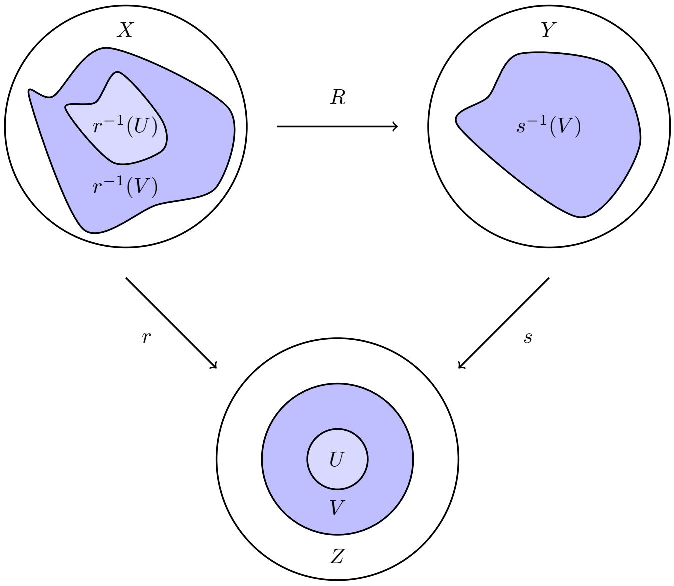


Figure 10.2. An admissible diagram.

LEMMA 10.7. *The sets* $\mathrm{Sing}(r) \smallsetminus U$ *and* $i^{-1}(\mathrm{Sing}(r)) \smallsetminus U$ *are mutually separated.*

PROOF. This will be a straightforward consequence of the fact that $P := i((\mathrm{Sing}(r) \smallsetminus U) \cap V)$ and $Q := \mathrm{Sing}(r) \smallsetminus U$ are mutually separated. Note that $\overline{P} = \overline{i(\mathrm{Sing}(r) \smallsetminus U \cap V)} = i(\overline{\mathrm{Sing}(r) \smallsetminus U} \cap V)$ since i is a homeomorphism and V is closed.

Suppose that

$$x \in (\mathrm{Sing}(r) \smallsetminus U) \cap \overline{i^{-1}(\mathrm{Sing}(r)) \smallsetminus U}.$$

Then since $x \in \overline{i^{-1}(\mathrm{Sing}(r)) \smallsetminus U}$, we know that $x \in V$. Moreover, we know that

$$\overline{i^{-1}(\mathrm{Sing}(r)) \smallsetminus U} = \overline{i^{-1}(\mathrm{Sing}(r) \smallsetminus U)} = i^{-1}(\overline{\mathrm{Sing}(r) \smallsetminus U}),$$

since i is a homeomorphism and restricts to the identity on U, so we see that $i(x) \in \overline{\mathrm{Sing}(r) \smallsetminus U} = \overline{Q}$. On the other hand, we know that $x \in (\mathrm{Sing}(r) \smallsetminus U) \cap V$ so $i(x) \in i((\mathrm{Sing}(r) \smallsetminus U) \cap V) = P$, which is a contradiction since P and Q are mutually separated. Thus $(\mathrm{Sing}(r) \smallsetminus U) \cap \overline{i^{-1}(\mathrm{Sing}(r)) \smallsetminus U} = \emptyset$.

Now suppose that

$$x \in \overline{(\mathrm{Sing}(r) \smallsetminus U)} \cap (i^{-1}(\mathrm{Sing}(r)) \smallsetminus U).$$

As before, since $x \in i^{-1}(\mathrm{Sing}(r)) \smallsetminus U$, we have that $x \in V$. Then, $x \in \overline{(\mathrm{Sing}(r) \smallsetminus U)} \cap V$ so $i(x) \in i(\overline{(\mathrm{Sing}(r) \smallsetminus U)} \cap V) = \overline{P}$. On the other hand, $x \in i^{-1}(\mathrm{Sing}(r)) \smallsetminus U$ so that $i(x) \in \mathrm{Sing}(r) \smallsetminus U = Q$, since i is a homeomorphism and acts by the identity on U, which is a contradiction since P and Q are mutually separated. It follows that $\overline{(\mathrm{Sing}(r) \smallsetminus U)} \cap (i^{-1}(\mathrm{Sing}(r)) \smallsetminus U) = \emptyset$. This completes the proof of Lemma 10.7. □

We return to the proof of Lemma 10.6. Note that $s^{-1} \circ i^{-1} \circ r \colon \partial X \to \partial(s^{-1}(V))$ is a homeomorphism since r is a homeomorphism on $\partial X \subseteq F$ and $\partial V \cap \mathrm{Sing}(s) = \emptyset$. The latter follows from the requirement that

$$V \subseteq Z \smallsetminus (\overline{\mathrm{Sing}(s)} \cup s(E)).$$

Apply the Alexander trick to produce a homeomorphism

$$j\colon X \to s^{-1}(V) \subseteq Y$$

such that $j = s^{-1} \circ i^{-1} \circ r$ on ∂X.

After making these choices, define the following relation $R_1\colon X \to Y$.

$$R_1 = \begin{cases} R & \text{on } X \smallsetminus r^{-1}(V) \\ j \circ (r^{-1} \circ i \circ r) & \text{on } r^{-1}(V \smallsetminus U) \\ j & \text{on } r^{-1}(U). \end{cases}$$

Note that this is roughly the same as the formula (10.1) for f_1; the only thing that is slightly different is the definition of j, which arises from the fact that we have three copies of D^4 in play.

Now we want a new admissible diagram using the relation R_1 just constructed. The relation R_1 is surjective by construction. Set $S_1 = R_1^{-1}$ and $r_1 = r$. As noted earlier, the pair (R_1, r_1) determines an admissible diagram $(R_1, S_1; r_1, s_1)$, where for s_1 we are forced to choose

$$s_1 = \begin{cases} s & \text{on } Y \smallsetminus s^{-1}(V) \\ i^{-1} \circ r \circ j^{-1} & \text{on } s^{-1}(V). \end{cases}$$

The above follows since we require $s_1 = r_1 \circ S_1 = r \circ R_1^{-1}$ and $i^{-1}(U) = U$. Note that $j^{-1} = r^{-1} \circ i \circ s$ on $s^{-1}(\partial V)$ so the formula is well defined on the overlap $s^{-1}(\partial V)$.

The next two lemmas verify that $(R_1, S_1; r_1, s_1)$ is indeed admissible.

LEMMA 10.8. *The relation R_1 is closed.*

PROOF. Observe that the relations R and $j \circ (r^{-1} \circ i \circ r)$ agree on $r^{-1}(\partial V)$, since j restricts to $s^{-1} \circ i^{-1} \circ r$ on ∂X, and the relations $j \circ (r^{-1} \circ i \circ r)$ and j agree on $r^{-1}(\partial U)$, since $i|_U$ is the identity map. Thus, the following is a valid definition of R_1.

$$R_1 = \begin{cases} R & \text{on } X \smallsetminus r^{-1}(\operatorname{Int} V) \\ j \circ (r^{-1} \circ i \circ r) & \text{on } r^{-1}(V \smallsetminus \operatorname{Int} U) \\ j & \text{on } r^{-1}(U). \end{cases}$$

Then it is clear that $R_1 \subseteq X \times Y$ is the union

$$\begin{aligned} &\{(x,y) \mid x \in X \smallsetminus r^{-1}(\operatorname{Int} V), (x,y) \in R\} \\ &\cup \{(x,y) \mid x \in r^{-1}(V \smallsetminus \operatorname{Int} U), y \in j \circ r^{-1} \circ i \circ r\{x\}\} \\ &\cup \{(x,y) \mid x \in r^{-1}(U), y = j(x)\}. \end{aligned}$$

We will show that each of the above sets is closed, which will establish that R_1 is closed.

First, note that

$$\{(x,y) \mid x \in X \smallsetminus r^{-1}(\operatorname{Int} V), (x,y) \in R\} = R \cap ((X \smallsetminus r^{-1}(\operatorname{Int} V)) \times Y).$$

The latter set is closed since R is closed by hypothesis, $\operatorname{Int} V$ is open, and r is continuous.

Similarly, $\{(x,y) \mid x \in r^{-1}(U), y = j(x)\}$ is the graph of the homeomorphism $j|_{r^{-1}(U)}\colon r^{-1}(U) \to j \circ r^{-1}(U)$ and is thus closed in $r^{-1}(U) \times (j \circ r^{-1}(U))$ by the closed graph lemma (Lemma 10.2). The latter set is closed in $X \times Y$, so the set $\{(x,y) \mid x \in r^{-1}(U), y = j(x)\}$ is closed in $X \times Y$.

It remains only to show that

$$\{(x,y) \mid x \in r^{-1}(V \smallsetminus \operatorname{Int} U), y \in j \circ r^{-1} \circ i \circ r\{x\}\},$$

is closed in $X \times Y$. This will be slightly more complicated. First we show that $r^{-1} \circ i \circ r \colon X \to X$ is a closed relation, i.e. the set

$$\{(x, x') \mid x' \in r^{-1} \circ i \circ r\{x\}\} \subseteq X \times X$$

is closed. Note that $i \circ r$ defines a closed set in $X \times Z$ since $i \circ r$ is a continuous function by the closed graph lemma (Lemma 10.2). Define the continuous function

$$\widetilde{r} \colon X \times X \to X \times Z$$

by setting $\widetilde{r}(x, x') = (x, r(x'))$. Then we see that

$$\{(x, x') \mid x' \in r^{-1} \circ i \circ r\{x\}\} = \widetilde{r}^{-1}\{(x, z) \mid z = i \circ r(x)\}$$

and is thus closed. Next, define the continuous function

$$\widetilde{j} \colon X \times X \to X \times s^{-1}(V)$$

given by $\widetilde{j}(x, x') = (x, j(x'))$. As a continuous map from a compact space to a Hausdorff space, $\widetilde{j}$ is closed by the closed map lemma (Lemma 3.23). Thus

$$\{(x, y) \mid y \in j \circ r^{-1} \circ i \circ r\{x\}\} = \widetilde{j}(\{(x, x') \mid x' \in r^{-1} \circ i \circ r\{x\}\})$$

is closed in $X \times s^{-1}(V)$. The latter is a closed subset of $X \times Y$, so the set

$$\{(x, y) \mid y \in j \circ r^{-1} \circ i \circ r\{x\}\}$$

is closed in $X \times Y$.

Note that the set

$$\{(x, y) \mid x \in r^{-1}(V \smallsetminus \operatorname{Int} U), y \in j \circ r^{-1} \circ i \circ r\{x\}\}$$

is equal to the intersection

$$\{(x, y) \mid y \in j \circ r^{-1} \circ i \circ r\{x\}\} \cap r^{-1}(V \smallsetminus \operatorname{Int} U) \times Y\},$$

which is closed in $X \times Y$ as the intersection of two closed sets. This finishes the proof of Lemma 10.8 that the relation R_1 is closed. □

LEMMA 10.9. *The quadruple $(R_1, S_1; r_1, s_1)$ is an admissible diagram.*

PROOF. We have already seen that $R_1 \colon X \to Y$ is closed and surjective, which implies that $S_1 := R_1^{-1}$ is closed and surjective. The function $r_1 := r$ is continuous by assumption. The function s_1 is continuous since the functions s and $i^{-1} \circ r \circ j^{-1}$ agree on the overlap $s^{-1}(\partial V)$ as already shown. The diagram commutes by construction. It is clear that R_1 and r_1 are homeomorphisms on F, and S_1 and s_1 are homeomorphisms on E by the choice of V. It remains to verify the conditions on the inverse sets and singular images.

By (10.2) and the choice of V, we have ensured that inverse sets do not get combined by accident. More precisely, we have that

$$\operatorname{Sing}(r_1) = \operatorname{Sing}(r)$$

and

$$\operatorname{Sing}(s_1) = \operatorname{Sing}(s) \cup i^{-1}(\operatorname{Sing}(r)).$$

This follows from the definition of s_1 and the fact that V does not contain any singular points of s. Thus, the inverse sets of s_1 are the inverse sets of s union the image of the inverse sets of r under j. Since j is uniformly continuous by the Heine-Cantor theorem (Theorem 3.25), the collection of inverse sets of s_1 is null since the inverse sets for s and r are null by hypothesis. Since nowhere density is preserved under homeomorphisms and finite union, we see that $\operatorname{Sing}(s_1)$ and $\operatorname{Sing}(r_1)$ are nowhere dense.

Next we show mutual separation. Recall that

$$\begin{aligned} A &= \operatorname{Sing}(r) \smallsetminus \operatorname{Sing}(s) \\ B &= \operatorname{Sing}(s) \smallsetminus \operatorname{Sing}(r) \\ C &= \operatorname{Sing} r \cap \operatorname{Sing}(s). \end{aligned}$$

We have the sets

$$\begin{aligned} A_1 &:= \operatorname{Sing}(r_1) \smallsetminus \operatorname{Sing}(s_1) \subseteq A \smallsetminus U, \\ B_1 &:= \operatorname{Sing}(s_1) \smallsetminus \operatorname{Sing}(r_1) = B \cup (i^{-1}(\operatorname{Sing}(r)) \smallsetminus U), \text{ and} \\ C_1 &:= \operatorname{Sing}(r_1) \cap \operatorname{Sing}(s_1) = C \cup (A \cap U). \end{aligned}$$

As an aside, note that as desired, the set A_1 is smaller than A and in particular, the singular point $a \in U$ is no longer in $\operatorname{Sing}(r)$ but now lies in $\operatorname{Sing}(s)$. Similarly, the sets B and C have increased in size as intended.

We claim that A_1, B_1, and C_1 are mutually separated. This will be straightforward (but involved) set manipulation. First we recall the facts we will need for the proof. Recall from Lemma 10.7 that $\operatorname{Sing}(r) \smallsetminus U$ and $i^{-1}(\operatorname{Sing}(r)) \smallsetminus U$ are mutually separated, that is

$$\begin{aligned} (\operatorname{Sing}(r) \smallsetminus U) \cap \overline{i^{-1}(\operatorname{Sing}(r)) \smallsetminus U} &= \emptyset \\ \overline{\operatorname{Sing}(r) \smallsetminus U} \cap i^{-1}(\operatorname{Sing}(r)) \smallsetminus U &= \emptyset. \end{aligned} \tag{10.3}$$

We will use that

$$A \cap \overline{B \cup C} = B \cap \overline{C \cup A} = C \cap \overline{A \cup B} = \emptyset \tag{10.4}$$

since A, B, and C are mutually separated. We will also use that

$$\operatorname{Sing}(r) \cap \partial U = \emptyset = \operatorname{Sing}(s) \cap U, \tag{10.5}$$

and that U is closed. Recall that for subsets K and L of some space we have $\overline{K \cup L} = \overline{K} \cup \overline{L}$ and $\overline{K \cap L} \subseteq \overline{K} \cap \overline{L}$. As usual we write K^c for the complement of K.

Now we begin the proof of the claim that

$$A_1 \cap \overline{B_1 \cup C_1} = B_1 \cap \overline{C_1 \cup A_1} = C_1 \cap \overline{A_1 \cup B_1} = \emptyset.$$

We have:

$$\begin{aligned} & A_1 \cap \overline{B_1 \cup C_1} \\ \subseteq & (A \smallsetminus U) \cap \overline{(B \cup (i^{-1}(\operatorname{Sing}(r)) \smallsetminus U) \cup C \cup (A \cap U))} \\ = & ((A \smallsetminus U) \cap \overline{B \cup C}) \cup \big((A \smallsetminus U) \cap \overline{i^{-1}(\operatorname{Sing}(r)) \smallsetminus U}\big) \\ & \cup (A \smallsetminus U \cap \overline{A \cap U}) \\ \subseteq & (A \cap \overline{B \cup C}) \cup \big((A \smallsetminus U) \cap \overline{i^{-1}(\operatorname{Sing}(r)) \smallsetminus U}\big) \cup (A \smallsetminus U \cap \overline{A} \cap \overline{U}) \\ = & \big((A \smallsetminus U) \cap \overline{i^{-1}(\operatorname{Sing}(r)) \smallsetminus U}\big) \cup (A \smallsetminus U \cap \overline{A} \cap U) \\ = & (A \smallsetminus U) \cap \overline{i^{-1}(\operatorname{Sing}(r)) \smallsetminus U} \\ \subseteq & (\operatorname{Sing}(r) \smallsetminus U) \cap \overline{i^{-1}(\operatorname{Sing}(r)) \smallsetminus U} \end{aligned}$$

which is empty by Lemma 10.7 i.e. the first equation of (10.3). We used that $A \cap \overline{B \cup C} = \emptyset$ by (10.4), and that $U = \overline{U}$ since U is closed. Next, we have:

$$\begin{aligned}
&B_1 \cap \overline{C_1 \cup A_1} \\
\subseteq&(B \cup (i^{-1}(\mathrm{Sing}(r)) \smallsetminus U)) \cap \overline{C \cup (A \cap U) \cup A \smallsetminus U} \\
=&(B \cup (i^{-1}(\mathrm{Sing}(r)) \smallsetminus U)) \cap \overline{C \cup A} \\
=&(B \cap \overline{C \cup A}) \cup ((i^{-1}(\mathrm{Sing}(r)) \smallsetminus U) \cap \overline{C \cup A}) \\
=&(i^{-1}(\mathrm{Sing}(r)) \smallsetminus U) \cap \overline{\mathrm{Sing}(r)} \\
=&(i^{-1}(\mathrm{Sing}(r)) \smallsetminus U) \cap \overline{\mathrm{Sing}(r) \smallsetminus U \cup (\mathrm{Sing}(r) \cap U)} \\
=&(i^{-1}(\mathrm{Sing}(r)) \smallsetminus U) \cap \big(\overline{\mathrm{Sing}(r) \smallsetminus U} \cup \overline{\mathrm{Sing}(r) \cap U}\big) \\
=&\big((i^{-1}(\mathrm{Sing}(r)) \smallsetminus U) \cap \overline{\mathrm{Sing}(r) \smallsetminus U}\big) \cup \big(i^{-1}(\mathrm{Sing}(r)) \smallsetminus U \cap \overline{\mathrm{Sing}(r) \cap U}\big) \\
\subseteq&\big((i^{-1}(\mathrm{Sing}(r)) \smallsetminus U) \cap \overline{\mathrm{Sing}(r) \smallsetminus U}\big) \cup \big(i^{-1}(\mathrm{Sing}(r)) \smallsetminus U \cap \overline{\mathrm{Sing}(r)} \cap \overline{U}\big) \\
=&\big((i^{-1}(\mathrm{Sing}(r)) \smallsetminus U) \cap \overline{\mathrm{Sing}(r) \smallsetminus U}\big) \cup \big(i^{-1}(\mathrm{Sing}(r)) \smallsetminus U \cap \overline{\mathrm{Sing}(r)} \cap U\big) \\
=&(i^{-1}(\mathrm{Sing}(r)) \smallsetminus U) \cap \overline{\mathrm{Sing}(r) \smallsetminus U}
\end{aligned}$$

which is empty by Lemma 10.7 i.e. the second equation of (10.3). We used that $B \cap \overline{C \cup A} = \emptyset$ by mutual separation (10.4), and we used again that $U = \overline{U}$. Then, for the last intersection that we must show is empty in order to prove the claim that A_1, B_1, and C_1 are mutually separated, we have:

$$\begin{aligned}
&C_1 \cap \overline{A_1 \cup B_1} \\
\subseteq&(C \cup (A \cap U)) \cap \overline{(A \smallsetminus U) \cup B \cup (i^{-1}(\mathrm{Sing}(r)) \smallsetminus U)} \\
=&\big(C \cap (\overline{A \smallsetminus U \cup B})\big) \cup \big(C \cap \overline{i^{-1}(\mathrm{Sing}(r)) \smallsetminus U}\big) \cup \big((A \cap U) \cap (\overline{A \smallsetminus U})\big) \\
&\cup \big(A \cap U \cap \overline{B}\big) \cup \big(A \cap U \cap \overline{i^{-1}(\mathrm{Sing}(r)) \smallsetminus U}\big).
\end{aligned}$$

Now $C \cap (\overline{A \smallsetminus U \cup B}) \subseteq C \cap (\overline{A \cup B}) = \emptyset$ and $A \cap U \cap \overline{B} \subseteq A \cap \overline{B} = \emptyset$ by mutual separation (10.4). Also we have:

$$\begin{aligned}
&(A \cap U) \cap \overline{A \smallsetminus U} \\
=&A \cap U \cap \overline{A \cap U^c} \\
\subseteq&A \cap U \cap \overline{A} \cap \overline{U^c} \\
=&A \cap \overline{A} \cap \partial U \\
=&A \cap \partial U \\
=&\mathrm{Sing}(r) \cap \partial U = \emptyset,
\end{aligned}$$

where the last two equalities use (10.5). Therefore $C_1 \cap \overline{A_1 \cup B_1}$ simplifies to

$$\begin{aligned}
&\big(C \cap \overline{i^{-1}(\mathrm{Sing}(r)) \smallsetminus U}\big) \cup \big(A \cap U \cap \overline{i^{-1}(\mathrm{Sing}(r)) \smallsetminus U}\big) \\
=&\big(\,\mathrm{Sing}(r) \cap \mathrm{Sing}(s) \cap \overline{i^{-1}(\mathrm{Sing}(r)) \smallsetminus U}\big) \\
&\cup \big(\,\mathrm{Sing}(r) \cap \mathrm{Sing}(s)^c \cap U \cap \overline{i^{-1}(\mathrm{Sing}(r)) \smallsetminus U}\big).
\end{aligned}$$

We show that this is empty for each of the components of the union separately. First, using that $\mathrm{Sing}(s) \cap U = \emptyset$ by (10.5) we have that:

$$\begin{aligned}
&\mathrm{Sing}(r) \cap \mathrm{Sing}(s) \cap \overline{i^{-1}(\mathrm{Sing}(r)) \smallsetminus U} \\
\subseteq&\,\mathrm{Sing}(r) \cap U^c \cap \overline{i^{-1}(\mathrm{Sing}(r)) \smallsetminus U} \\
=&(\mathrm{Sing}(r) \smallsetminus U) \cap \overline{i^{-1}(\mathrm{Sing}(r)) \smallsetminus U}
\end{aligned}$$

which is empty by Lemma 10.7 i.e. (10.3). Second, we have:

$$\begin{aligned}
&\mathrm{Sing}(r)\cap \mathrm{Sing}(s)^c \cap U \cap \overline{i^{-1}(\mathrm{Sing}(r)) \smallsetminus U} \\
&= \mathrm{Sing}(r)\cap \mathrm{Sing}(s)^c \cap U \cap \overline{i^{-1}(\mathrm{Sing}(r)) \cap U^c} \\
&\subseteq \mathrm{Sing}(r)\cap \mathrm{Sing}(s)^c \cap U \cap \overline{i^{-1}(\mathrm{Sing}(r))} \cap \overline{U^c} \\
&= \mathrm{Sing}(r)\cap \mathrm{Sing}(s)^c \cap \partial U \cap \overline{i^{-1}(\mathrm{Sing}(r))}
\end{aligned}$$

which is empty since $\mathrm{Sing}(r)\cap\partial U = \emptyset$ by (10.5). This completes the demonstration that $C_1 \cap \overline{A_1 \cup B_1} = \emptyset$. We have shown the claim that the sets A_1, B_1, and C_1 are mutually separated.

Finally, since

$$R_1 = \begin{cases} R & \text{on } r^{-1}(C) \\ j & \text{on } r^{-1}(A\cap U), \end{cases}$$

by definition, since $C \subseteq Z \smallsetminus V$, we see that $R_1 \colon r_1^{-1}(C_1) \to s_1^{-1}(C_1)$ is a homeomorphism as desired. This completes the proof of Lemma 10.9. □

We are almost finished with the proof of Lemma 10.6. Choose V small enough (also making U smaller, as needed) to guarantee that R_1 is contained in $\mathrm{Int}\,\mathcal{N}(R)$. This is possible since R_1 differs from R only in $r^{-1}(V)\times s^{-1}(V)$, so we choose V small enough that $r^{-1}(V)\times s^{-1}(V) \subseteq \mathrm{Int}\,\mathcal{N}(R)$.

Then repeat the construction finitely many times, once for each element $a \in A^\varepsilon$, to obtain relations $R_2, R_3, \ldots$ and the corresponding admissible diagrams for the rest of the points in A^ε. Let $(R' \colon X \to Y, S'; r', s')$ be the final relation obtained in this way. By construction, we have that $r' = r$, $R'|_F = R|_F = f$, and $R' \in \mathcal{N}(R)$. Note that $A' := \mathrm{Sing}(r') \smallsetminus \mathrm{Sing}(s') \subseteq A \smallsetminus A^\varepsilon$. In particular, we have removed all the singular image points coming from inverse sets with diameter greater than ε, but we may have fortuitously removed some others as well. Since $s' \circ \mathrm{Sing}(R') = A'$ (Lemma 10.5), note that we have eliminated all the inverse sets of R of diameter at least ε. Roughly speaking, the large inverse sets of R correspond to elements of C' and we have arranged that R' is a homeomorphism when restricted to ${r'}^{-1}(C')$. In other words, $\mathrm{hd}(R') < \varepsilon$. This completes the proof of Lemma 10.6. □

10.4. Proof of the ball to ball theorem

Let X and Y denote two copies of D^4. We start with the continuous function $f \colon X \to Y$, which gives a closed subset of $X \times Y$ by Lemma 10.2. Choose a closed neighbourhood $\mathcal{N}(f) \subseteq X\times Y$ with $f \subseteq \mathrm{Int}\,\mathcal{N}(f)$. Our goal for the proof of Theorem 10.1 is to construct a homeomorphism $h\colon X \to Y$ close to f, meaning that $h \subseteq \mathcal{N}(f)$. This will imply that f is approximable by homeomorphisms, directly from Definition 3.19.

The idea is to construct a sequence $\{R_n\}_{n\geq 1}$ of closed relations in $\mathcal{N}(f)$, with decreasing horizontal and vertical defects, and take the limit. Lemma 10.6 allows us to make the horizontal defect arbitrarily small, while the vertical defect may increase a little bit, being controlled by the given neighbourhood of the original relation. In order to gain more control on the vertical defect, we shall take advantage of the symmetry of our situation and apply Lemma 10.6 on $S_n = R_n^{-1}$ as well. In this way, we can reduce, alternately, both the horizontal and the vertical defects. We shall obtain the homeomorphism h as the intersection of a nested sequence of compact sets in $\mathcal{N}(f)$, instead of considering the limit of a sequence of relations.

Now we give the formal construction. Let $\{\varepsilon_n\}_{n\geq 1}$ be a positive monotone decreasing sequence which converges to zero. We will define a sequence of admissible diagrams $\{\mathcal{A}_n\}_{n\geq 1}$, where $\mathcal{A}_n = (R_n, S_n; r_n, s_n)$ for $n \geq 1$. By the assumptions, $(f, f^{-1}; f, \mathrm{Id})$ is an admissible diagram. Here we are using that f^{-1} is closed since f is closed. Apply Lemma 10.6 on $(f, f^{-1}; f, \mathrm{Id})$ with $\varepsilon = \varepsilon_1$, and $\mathcal{N}(R) = \mathcal{N}(f)$ to obtain an admissible diagram $\mathcal{A}_1 = (R_1, S_1; r_1, s_1)$ such that $\mathrm{hd}(R_1) < \varepsilon_1$ and $R_1 \subseteq \operatorname{Int} \mathcal{N}(f)$. Whenever we apply Lemma 10.6, the sets E and F needed for the input will be the sets E and $f^{-1}(E)$ from the hypotheses of the ball to ball theorem, in some order.

Choose a closed neighbourhood $\mathcal{D}_1$ of R_1 in $X \times Y$ such that $\mathcal{D}_1 \subseteq \mathcal{N}(f)$, $R_1 \subseteq \operatorname{Int} \mathcal{D}_1$, and $\mathrm{hd}(\mathcal{D}_1) < \varepsilon_1$. One may find such a $\mathcal{D}_1$ by taking the intersection of the closed set given by the following lemma with $\mathcal{N}(f)$.

Lemma 10.10 ([**Anc84**, Lemma 2]). *Let X and Y be compact metric spaces and $\varepsilon > 0$. Let $R \subseteq X \times Y$ be closed. Assume that $\mathrm{hd}(R) < \varepsilon$. Then there exists a closed set $D \subseteq X \times Y$ with $R \subseteq \operatorname{Int} D$ and $\mathrm{hd}(D) < \varepsilon$.*

Proof. We define the following notation for use in this proof. Let $A \subseteq Y$ and $Q \subseteq X \times Y$. Then we define $Q|_A := \{(x, y) \in Q \mid y \in A\}$.

Let $\{C_i\}$ be a sequence of closed neighbourhoods of $X \times Y$ with $R \subseteq C_i$ for each i and $\bigcap C_i = R$. Then for any compact set $A \subseteq Y$, there exists some i so that $\mathrm{hd}(C_i|_A) < \varepsilon$. Suppose not. Then there exist sequences $\{(x_i, y_i)\}$ and $\{(x_i', y_i)\}$ in $X \times A$ with $d(x_i, x_i') \geq \varepsilon$ for all i. Since X and A are compact, we can assume, after passing to subsequences, that $\{x_i\}$, $\{x_i'\}$, and $\{y_i\}$ converge to x, x', and y' respectively. Then $(x, y), (x', y) \in R$ but $d(x, x') \geq \varepsilon$ which is a contradiction since $\mathrm{hd}(R) < \varepsilon$.

Next let $\{A_i\}$ be a sequence of compact subsets of Y with $A_i \subseteq \operatorname{Int} A_{i+1}$ for all i with $\bigcup A_i = Y$. Such an *exhaustion by compact sets* exists since Y is a compact metric space [**Lee11**, Proposition 4.76]. By the argument above, it follows that we may pass to a subsequence of $\{C_i\}$ to obtain a sequence $\{D_i\}$ of closed neighbourhoods of R with $\bigcap D_i = R$ and for all i, $\mathrm{hd}(D_i|_{A_i}) < \varepsilon$. Then set $D = \bigcup_{i\geq 1} D_i|_{A_i}$. Define $P_i = \bigcup_{j=1}^{i-1} D_j|_{A_i} \cup D_i$. Note that each P_i is closed, so $D = \bigcap_{i\geq 1} P_i$ is also closed. For each i, we have that $P_i|_{A_i} = D|_{A_i}$, so $R|_{\operatorname{Int} A_i} \subseteq \operatorname{Int} P_i|_{\operatorname{Int} A_i} \subseteq \operatorname{Int} D$. It follows that $R|_Y = R \subseteq \operatorname{Int} D$ as needed. For each i, if $y \in A_i \smallsetminus A_{i-1}$, we know that $\operatorname{diam}_X D^{-1}\{y\} < \varepsilon$ since $D^{-1}\{y\} = D_i^{-1}\{y\}$ and we know that $\mathrm{hd}(D_i) < \varepsilon$. □

Next apply Lemma 10.6 on $(S_1, R_1; s_1, r_1)$, $\varepsilon = \varepsilon_1$, and $\mathcal{N}(S_1) = \mathcal{D}_1^{-1}$, to obtain an admissible diagram $(S_2, R_2; s_2, r_2)$ such that $S_2 \subseteq \operatorname{Int} \mathcal{D}_1^{-1}$ and $\mathrm{vd}(R_2) = \mathrm{hd}(S_2) < \varepsilon_1$. The admissible diagram $\mathcal{A}_2$ is defined to be $(R_2, S_2; r_2, s_2)$. Use the Lemma 10.10 to choose a closed neighbourhood $\mathcal{D}_2$ of S_2 such that $\mathcal{D}_2 \subseteq \mathcal{D}_1^{-1}$ and $\mathrm{hd}(\mathcal{D}_2) < \varepsilon_1$.

In general, suppose that the admissible diagram $\mathcal{A}_{2k} = (R_{2k}, S_{2k}; r_{2k}, s_{2k})$ has been defined and a closed neighbourhood $\mathcal{D}_{2k}$ of S_{2k} has been chosen satisfying $\mathcal{D}_{2k} \subseteq \mathcal{D}_{2k-1}^{-1}$ and $\mathrm{hd}(\mathcal{D}_{2k}) < \varepsilon_k$. Apply Lemma 10.6 to $(R_{2k}, S_{2k}; r_{2k}, s_{2k})$, $\mathcal{N}(R_{2k}) = \mathcal{D}_{2k}^{-1}$, and $\varepsilon = \varepsilon_{k+1}$ to obtain an admissible diagram

$$\mathcal{A}_{2k+1} := (R_{2k+1}, S_{2k+1}; r_{2k+1}, s_{2k+1}).$$

Use Lemma 10.10 to choose a closed neighbourhood $\mathcal{D}_{2k+1}$ of R_{2k+1} such that $\mathcal{D}_{2k+1} \subseteq \mathcal{D}_{2k}^{-1}$, and $\mathrm{hd}(\mathcal{D}_{2k+1}) < \varepsilon_{k+1}$. This gives rise to an admissible diagram $(S_{2k+2}, R_{2k+2}; s_{2k+2}, r_{2k+2})$ with $\mathrm{hd}(S_{2k+2}) < \varepsilon_{k+1}$ and $S_{2k+2} \subseteq \operatorname{Int} \mathcal{D}_{2k+1}^{-1}$ by applying Lemma 10.6 to the 4-tuple $(S_{2k+1}, R_{2k+1}; s_{2k+1}, r_{2k+1})$, $\mathcal{N}(S_{2k+1}) = \mathcal{D}_{2k+1}^{-1}$, and $\varepsilon = \varepsilon_{k+1}$. Then define

$$\mathcal{A}_{2k+2} := (R_{2k+2}, S_{2k+2}; r_{2k+2}, s_{2k+2}).$$

Use Lemma 10.10 to choose a closed neighbourhood $\mathcal{D}_{2k+2}$ of S_{2k+2} such that $\mathcal{D}_{2k+2} \subseteq \mathcal{D}_{2k+1}^{-1}$ and $\mathrm{hd}(\mathcal{D}_{2k+2}) < \varepsilon_{k+1}$.

This process yields the relations R_{2n-1}, S_{2n}, and $\mathcal{D}_n$ for every $n \geq 1$. By construction, the closed relations $\mathcal{D}_n$ satisfy $\mathcal{D}_{n+1} \subseteq \mathcal{D}_n^{-1}$, $R_{2n-1} \subseteq \operatorname{Int} \mathcal{D}_{2n-1}$, and $S_{2n} \subseteq \operatorname{Int} \mathcal{D}_{2n}$. Moreover, $\{\mathrm{hd}(\mathcal{D}_i)\}$ and $\{\mathrm{vd}(\mathcal{D}_i)\}$ converge to zero. Let h be the intersection of the following chain of closed sets

$$\cdots \subseteq \mathcal{D}_{2n+2}^{-1} \subseteq \mathcal{D}_{2n+1} \subseteq \mathcal{D}_{2n}^{-1} \subseteq \mathcal{D}_{2n-1} \subseteq \mathcal{D}_{2n-2}^{-1} \subseteq \cdots .$$

The intersection is nonempty (see Lemma 7.1). It follows immediately that h is a continuous injective function, since $\mathrm{hd}(h) = \mathrm{vd}(h) = 0$ and h is closed by construction. The domain of h is all of X because the surjectivity of S_{2n} implies that $S_{2n}^{-1}\{x\}$ is nonempty for every $x \in X$. Also the image of h is all of Y because the surjectivity of R_{2n-1} implies that $R_{2n-1}^{-1}\{y\}$ is nonempty for every $y \in Y$. Since h is a bijective map between compact metric spaces it is a homeomorphism by the closed map lemma (Lemma 3.23). This concludes the proof of the ball to ball theorem (Theorem 10.1) modulo the general position lemma (Lemma 10.12), which we prove in the next section.

10.5. The general position lemma

In Section 4.1, we used the Baire category theorem (Theorem 4.2) to show the Bing shrinking criterion. We use the Baire category theorem again now to prove the general position lemma that was used in the proof of Lemma 10.6. A similar general position lemma is stated in [**Fre82a**, Lemma 9.1], but left as an exercise. We follow the exposition of Ancel [**Anc84**, Section 2], where a stronger version is proven in detail.

Recall that for compact metric spaces Z and Y, the complete metric space $\mathcal{C}(Z,Y)$ consists of continuous maps from Z to Y with the uniform metric

$$d(f,g) = \sup_{z \in Z} d_Y(f(z), g(z)).$$

Consider the set $\operatorname{Homeo}(Z) \subseteq \mathcal{C}(Z,Z)$ of self-homeomorphisms of Z. The uniform metric does not in general induce a complete metric on $\operatorname{Homeo}(Z)$. Instead, we define a complete metric ρ on $\operatorname{Homeo}(Z)$ by

$$\rho(f,g) := d(f,g) + d(f^{-1}, g^{-1}).$$

PROPOSITION 10.11. *For a compact metric space Z, the set* $\operatorname{Homeo}(Z)$ *equipped with the metric ρ is a complete metric space.*

PROOF. Let $\{f_i\}$ be a Cauchy sequence in $\operatorname{Homeo}(Z)$. Then since $d(f_n, f_m) < \rho(f_n, f_m)$ and $d(f_n^{-1}, f_m^{-1}) < \rho(f_n, f_m)$ for all n, m, the sequences $\{f_i\}$ and $\{f_i^{-1}\}$ are Cauchy in $\mathcal{C}(Z,Z)$. Let f and g denote the limits of these sequences respectively. Since each f_i and f_i^{-1} is uniformly continuous by the Heine-Cantor theorem (Theorem 3.25), and since taking limits of sequences of uniformly continuous functions commutes with composition (that is, if $\{f_i \colon Z \to Y\}$ has limit f and $\{g_j \colon Y \to X\}$ has limit g then $\{g_k \circ f_k \colon Z \to X\}$ has limit $f \circ g$), we have

$$f \circ g = \lim_i f_i \circ \lim_j f_j^{-1} = \lim_k (f_k \circ f_k^{-1}) = \lim \mathrm{Id} = \mathrm{Id} \text{ and}$$
$$g \circ f = \lim_i f_i^{-1} \circ \lim_j f_j = \lim_k (f_k^{-1} \circ f_k) = \lim \mathrm{Id} = \mathrm{Id}.$$

Therefore f and g are inverse to each other and so lie in $\operatorname{Homeo}(Z)$. □

Given any subset A of Z, we consider

$$\operatorname{Homeo}_A(Z) := \{f \in \operatorname{Homeo}(Z) \mid f|_A = \mathrm{Id}_A\},$$

equipped with the metric induced from ρ. Then observe that $\mathrm{Homeo}_A(Z)$ is a closed subset of $\mathrm{Homeo}(Z)$ since it contains all its limit points (in other words, if $\{f_i\}$ is a sequence with $f_i(a) = a$ for every $a \in A$, then the limit homeomorphism f also satisfies $f(a) = a$), and thus it is itself a complete metric space.

LEMMA 10.12 (General position lemma [**Anc84**, Lemma 1(2)]). *Let Z be a compact manifold with a metric d inducing its topology and let G be a closed subset. Let $P, Q \subseteq Z \smallsetminus G$ be countable and nowhere dense subsets in the complement of G. For every $\eta > 0$ there exists a self-homeomorphism h_η of Z, such that $d(h_\eta, \mathrm{Id}_Z) < \eta$, $h_\eta = \mathrm{Id}$ on G, and such that $h_\eta(P)$ and Q are mutually separated.*

PROOF. Write $P = \{c_i\}_{i=1}^\infty$ and $Q = \{d_i\}_{i=1}^\infty$. Define the sets

$$U_i = \{h \in \mathrm{Homeo}_G(Z) \mid h(c_i) \notin \overline{Q}\}$$
$$V_i = \{h \in \mathrm{Homeo}_G(Z) \mid d_i \notin \overline{h(P)}\}.$$

We claim that each U_i and V_i is open and dense in $\mathrm{Homeo}_G(Z)$.

To prove that U_i is open, define continuous functions $q_i\colon \mathrm{Homeo}_G(Z) \to Z$ by setting $q_i(f) = f(c_i)$. Then $U_i = q_i^{-1}(Z \smallsetminus \overline{Q})$ and is thus open.

To prove that U_i is dense, let $f \in \mathrm{Homeo}_G(Z)$ with $f(c_i) \in \overline{Q}$. Then for every $\eta > 0$, we wish to find $h \in U_i$ such that $\rho(f, h) < \eta$. Since f^{-1} is uniformly continuous by the Heine-Cantor theorem (Theorem 3.25), there is some δ such that $d(y, y') < \delta$ implies that $d(f^{-1}(y), f^{-1}(y')) < \eta/2$. Let B be the open ball of diameter $\min\{\eta/2, \delta\}$ around $f(c_i)$. By taking a smaller radius if necessary, ensure that B lies in the complement of G. Then the restriction $f|_{f^{-1}(B)}\colon f^{-1}(B) \to B$ is a map between balls and can be perturbed to a homeomorphism h which coincides with f on the boundary and $h(c_i) \notin \overline{Q}$. This uses that Q is nowhere dense and therefore $\overline{Q}$ has empty interior. Extend h by f over the complement of $f^{-1}(B)$ to produce $h \in \mathrm{Homeo}_G(Z)$. We need to show that $\rho(f, h) < \eta$. This follows since $d(f, h) < \eta/2$ and $d(f^{-1}, h^{-1}) < \eta/2$ by construction. This finishes the proof that U_i is dense in $\mathrm{Homeo}_G(Z)$ for each i.

Directly analogous arguments prove that each V_i is open and dense in $\mathrm{Homeo}_G(Z)$. This uses the fact that nowhere density is preserved under homeomorphism so $h(P)$ is nowhere dense.

We now use the Baire category theorem (Theorem 4.2). Recall that $\mathrm{Homeo}_G(Z)$ is a complete metric space by Proposition 10.11. Since the intersection of two open dense sets is open and dense, by the Baire category theorem $U := \bigcap U_i \cap V_i$ is dense in $\mathrm{Homeo}_G(Z)$. In other words, for every given η, there is a homeomorphism $h_\eta \in U$ so that $\rho(h_\eta, \mathrm{Id}_Z) < \eta$. By the definition of U, $h_\eta(P)$ and Q are mutually separated. We also see that $d(h_\eta, \mathrm{Id}_Z) < \eta$ by the straightforward inequality

$$d(h_\eta, \mathrm{Id}_Z) \leq \rho(h_\eta, \mathrm{Id}_Z) = d(h_\eta, \mathrm{Id}_Z) + d(h_\eta^{-1}, \mathrm{Id}_Z^{-1}) = \rho(h_\eta, \mathrm{Id}_Z) < \eta.$$

This completes the proof of the general position lemma. □

10.6. The sphere to sphere theorem from the ball to ball theorem

The original proof of the disc embedding theorem in [**Fre82a**] uses the sphere to sphere theorem. For completeness, following Ancel [**Anc84**, p. 148], we prove the sphere to sphere theorem from the ball to ball theorem. We will not use the sphere to sphere theorem in our proof of the disc embedding theorem.

THEOREM 10.13 (Sphere to sphere theorem). *Let $f\colon S^4 \to S^4$ be a surjective map such that the following hold.*

(*a*) *The collection of inverse sets is null.*
(*b*) *The singular image of f is nowhere dense.*

Then f can be approximated by homeomorphisms.

Proof. Let $\varepsilon > 0$. Since the singular image of f is nowhere dense, we can choose an embedded 4-ball B in S^4 such that $\operatorname{diam} B < \varepsilon$, B is disjoint from the singular image of f, and ∂B is bicollared. By the Schoenflies theorem (Theorem 3.29) applied to the bicollared 3-sphere $\partial B \subseteq S^4$, there exist homeomorphisms $\varphi\colon D^4 \to \overline{S^4 \smallsetminus f^{-1}(B)}$ and $\psi\colon D^4 \to \overline{S^4 \smallsetminus B}$. Since ψ is uniformly continuous by the Heine-Cantor theorem (Theorem 3.25), there is some $\delta > 0$ such that ψ sends any set of diameter less than δ to a set of diameter less than ε.

Consider the map $g := \psi^{-1} \circ f \circ \varphi\colon D^4 \to D^4$. Since ψ^{-1} is uniformly continuous by the Heine-Cantor theorem (Theorem 3.25), nullity is preserved under uniformly continuous maps, and nowhere density is preserved under homeomorphisms, the function g satisfies the conditions of Theorem 10.1 with $E = \partial D^4$. By the ball to ball theorem, we have a homeomorphism $g_\varepsilon\colon D^4 \to D^4$ such that g_ε is within δ of g, and g_ε coincides with g on ∂D^4.

Since B is disjoint from the singular image of f, the function

$$\psi \circ g_\varepsilon \circ \varphi^{-1}\colon \overline{S^4 \smallsetminus f^{-1}(B)} \to \overline{S^4 \smallsetminus B}$$

extends to a homeomorphism $f_\varepsilon\colon S^4 \to S^4$ since $\psi \circ g_\varepsilon \circ \varphi^{-1}$ coincides with f on $f^{-1}(\partial B)$. Note that f_ε is within ε of f by our choice of δ. $\square$

We close with some final remarks on the ball to ball theorem. First, many previous shrinking results showed that a quotient map is approximable by homeomorphisms, such as the quotient map $S^3 \to S^3/\mathcal{B}$ for the Bing decomposition in Theorem 5.4. But prior to Bing's proof, it was not even clear whether any homeomorphism between S^3 and $S^3/\mathcal{B}$ existed, let alone whether such homeomorphisms might approximate the quotient map. By contrast, the ball to ball theorem is applied to a decomposition whose decomposition space we already know to be a copy of D^4. The extra information provided by the theorem is the existence of a homeomorphism restricting to a predetermined homeomorphism $f^{-1}(E) \to E$, where E and $f^{-1}(E)$ are closed neighbourhoods of ∂D^4, and that is what we shall later use in the proof of the disc embedding theorem in Part IV.

While Part II contains many detailed constructions, these had to some extent been foreshadowed by Casson's work. The decomposition space theory from Part I appears in several places in the proof of the disc embedding theorem. But, as we have explained, the results of Chapters 5 to 9 are natural variations on results that, in the late 1970s, were well known in the geometric topology community. On the other hand, the ball to ball theorem, and the use of relations in its proof, was a radical departure from the methods previously employed. Consequently, this result of Freedman was reportedly the biggest shock to many experts in decomposition space theory. From this point of view, the proof we just gave is arguably the technical heart of the proof of the disc embedding theorem. It will appear again right at the end of Chapter 28.

This concludes our foray into the world of decomposition space theory.

Part II

Building skyscrapers

As mentioned in Chapters 1 and 2, the proof of the disc embedding theorem consists of building an infinite iterated object, called a *skyscraper*, and then showing that every skyscraper is homeomorphic to the standard 2-handle relative to its attaching region. In this part we perform the first step, constructing skyscrapers. The second step will take place in Part IV.

This part of the book will consist, after we give some key definitions, of hands on constructions in an ambient 4-dimensional space. In Chapter 11 we begin with a discussion of *intersection numbers*, leading to a careful statement of the disc embedded theorem. In Chapter 12 we define skyscrapers and their finite truncations, namely *gropes*, *capped gropes*, *towers*, and *capped towers*. These can be described precisely by the techniques of *Kirby calculus* using diagrams of links, as we discuss in Chapter 13. In Chapter 14 we collect the properties of skyscrapers and their finite truncations which will be useful to us in Part IV. Then in Chapter 15, we introduce the geometric objects and tools that we will use in the rest of this part.

In Chapter 16 we begin the proof of the disc embedding theorem in earnest. In particular, we replace the immersed discs in the hypotheses of the disc embedding theorem by a family of new immersed discs, with geometrically transverse spheres, with the same framed boundary as the original collection, whose intersections and self-intersections are paired by Whitney circles which bound height two capped gropes with geometrically transverse capped surfaces. In Chapter 17 we prove *grope height raising* and show how to upgrade these height two gropes to 1-storey capped towers with geometrically transverse spheres. In Chapter 18 we show that any 1-storey capped tower with at least four surface stages contains a skyscraper with the same attaching region. We finish this part by proving the disc embedding theorem, modulo the result from Part IV that skyscrapers are standard, in Section 18.4. In Chapter 18 we also prove an embedding property of skyscrapers, namely the skyscraper embedding theorem, which will be a key ingredient in Part IV.

CHAPTER 11

Intersection numbers and the statement of the disc embedding theorem

Mark Powell and Arunima Ray

We carefully state the disc embedding theorem, defining each term that appears therein. In particular we carefully describe intersection numbers.

11.1. Immersions

Let M be a smooth 4-manifold. Recall that the disc embedding theorem begins with smooth immersions and yields topological flat embeddings. As usual, an immersion is a local embedding and every map from a surface to a smooth 4-manifold can be approximated by an immersion. That is, every such function is homotopic to an immersion, which can be chosen to be arbitrarily close to the original function. If the boundary of the surface is already immersed, we may assume that it is fixed by the homotopy. We may assume moreover that the immersion is in general position, that is all intersections are transversal, lie in the interior, and are at most double points. Henceforth, we replace maps of surfaces in 4-manifolds by immersions and assume without comment that immersed surfaces have transverse double point intersections, all of which lie in the interior, and no triple points.

A *framed immersion* of an orientable surface F in M is an immersion of F in M such that the normal bundle of the image of F is trivial. Framed immersions can also be defined topologically as follows [**FQ90**, Section 1.2]. Given an abstract surface F, form the product $F \times \mathbb{R}^2$. Consider disjoint copies D and E of $\mathbb{R}^2$ in F. Perform a *plumbing* operation on $D \times \mathbb{R}^2$ and $E \times \mathbb{R}^2$. That is, identify $(x, y) \in D \times \mathbb{R}^2$ with $(y, x) \in E \times \mathbb{R}^2$. The standard orientations of D and E need not be restrictions of the same orientation on F, but we do require that the resulting 4-manifold be orientable. Repeated applications of this procedure for mutually disjoint D and E, yields a *plumbed model* for F. A (topological) *framed immersion* of the abstract surface F in M is a map from a plumbed model for F to some open set in M, that is a homeomorphism onto its image. Such a homeomorphism determines a map $g\colon F \to M$ when we restrict to the image of F in the framed model and we say that the map g *extends to a framed immersion*.

The normal bundle of a smoothly immersed, connected, orientable surface is determined up to isomorphism by its Euler number. If the Euler number is even, then the surface is homotopic via local cusp homotopies [**FQ90**, Section 1.6] to an immersion that extends to a framed immersion. In particular, a cusp homotopy changes the Euler number of the normal bundle of F by ± 2. Performing a local cusp homotopy was called adding a local kink in Chapter 1 (see Figure 1.3).

Since D^2 is contractible, the normal bundle of an immersed disc in M is trivial. Fixing an orientation of the fibres determines a framing of the normal bundle, uniquely up to homotopy, again since D^2 is contractible. Recall that a *framing* of a rank n vector bundle over a space B is by definition a trivialisation, namely an identification of the total space with $B \times \mathbb{R}^n$. So for $D^2 \looparrowright M$, this means an identification of the total space of the normal bundle with $D^2 \times \mathbb{R}^2$. If M and D^2 are both oriented, then the fibres of the normal bundle of D^2 inherit an orientation. But in general orientations of the fibres are an auxiliary choice that we make in order to proceed as in the next paragraph.

The framing of the normal bundle of an immersed disc induces a framing of the normal bundle restricted to the boundary ∂D^2. When a framing of this restricted normal bundle inducing the same orientation on fibres is independently specified, we may consider the *twisting number*, or *relative Euler number* of the induced framing with respect to the specified framing. This twisting number is an integer (coming from $\pi_1(SO(2)) \cong \mathbb{Z}$). In such a situation, we say that the immersed disc is *framed* when the twisting number is zero, so that the two framings match up to homotopy. For us, this situation will mostly occur in three ways:

(1) when the disc is *properly immersed*, that is when the preimage of ∂M is the boundary of the disc;
(2) when the disc is attached to a simple closed curve in an immersed surface in M, in which case we will consider the framing induced by the tangent bundle of the surface; or
(3) when the disc is a Whitney disc, in which case we consider the Whitney framing. This latter case will be described in more detail soon.

In general, given an immersed surface F in M, if the boundary of F is nonempty we will usually have a fixed framing already prescribed on the boundary. In this case we say that a map $g\colon F \to M$ *extends to a framed immersion* if it extends to a framed immersion restricting to the given framing on ∂F. For any immersed, connected surface F, with nonempty boundary $\partial F \subseteq M \smallsetminus \partial M$ in the interior of M, there is a homotopy via boundary twists, as described in Section 15.2.2 (see also [**FQ90**, Section 1.3]), to an immersion that extends to a framed immersion. This is because, as we will see, a single boundary twist changes the relative Euler number by ± 1.

A *regular homotopy* in the smooth category is a homotopy through immersions. It is well known that a smooth regular homotopy of immersed surfaces in a 4-manifold is generically a concatenation of smooth isotopies, fingers moves, and Whitney moves with respect to smoothly embedded and framed Whitney discs whose interiors are in the complement of the surfaces. Indeed, by [**GG73**, Section III.3], generic immersions are dense in the space of smooth mappings, and the generic singularities for a regular homotopy between surfaces are precisely arcs of isolated double points, which may appear (finger move) or disappear (Whitney move) at finitely many distinct time values. Whitney and finger moves, which are inverse to each other, were introduced in Chapter 1, and we give further details in a moment (see also [**FQ90**, Chapter 1]).

A *topological regular homotopy* of immersed surfaces in a 4-manifold is by definition a concatenation of (topological) isotopies, finger moves, and Whitney moves with respect to topologically flat embedded and framed Whitney discs whose interiors are in the complement of the surfaces. Regular homotopies of immersed surfaces with boundary take place in the interior of the surfaces, unless stated otherwise. For example, a finger move pushing an intersection point off the boundary of a disc, such as in Figure 11.4, is not permitted as part of a regular homotopy. Having said that, we will often change surfaces by this move (see for example Section 15.2.4). In that case, we perform an isotopy of the surface that moved, and a homotopy of the collection, but the motion does not count as a regular homotopy of the collection.

11.2. Whitney moves and finger moves

11.2.1. Whitney moves. Recall from Chapter 1 that a Whitney move is designed to remove two points of intersection between two immersed surfaces, or of an immersed surface with itself, within a smooth ambient 4-manifold M. If the associated Whitney disc is framed and embedded, with interior in the complement of the surfaces, then the resulting surfaces indeed have two fewer intersections. If the Whitney disc is framed but not embedded, or the interior intersects the surfaces, then the two original intersection points are still removed, but other double points might be introduced in the process. We now give further details.

Let f and g be two immersed oriented surfaces in a smooth 4-manifold M, with possibly $f = g$. Let p and q be two points in $f \pitchfork g$, such that there is an embedded arc γ in f from p to q and an embedded arc δ in g from p to q where the union $\gamma\delta^{-1}$ bounds an embedded disc D whose interior lies in the complement of f and g. See Figure 11.1. Fix a local orientation of M at p, and transport it along γ to q. Now, comparing with the orientations of T_pM and T_qM determined by the orientations of f and g yields a function $\operatorname{sgn}\colon \{p,q\} \to \{+,-\}$.

As before, the normal bundle of D in M is a trivial 2-plane bundle. Fix an orientation on the fibres. Consider the following 1-plane sub-bundle V of the normal bundle of D restricted to $\partial D = \gamma\delta^{-1}$. The sub-bundle along γ is given by the tangent bundle to f. This can be extended to a choice of sub-bundle along δ that is normal to g and agrees with Tf at p and q, since the intersections are transverse. This is a trivial 1-plane bundle if and only if the function $\operatorname{sgn}\colon \{p,q\} \to \{+,-\}$ is surjective, that is if and only if the signs of p and q are opposite.

Assuming that this is the case, choose a section s of the sub-bundle V. Since V is 1-dimensional, the section s is determined up to multiplication by a continuous function $S^1 \to \mathbb{R} \smallsetminus \{0\}$. We say that the Whitney disc D is *framed* if the section s extends to a nonvanishing section on the normal bundle of all of D. The framing of the normal bundle of D restricted to ∂D, induced by s and the chosen orientation on the fibres of the normal bundle, is called the *Whitney framing*.

Now extend the Whitney disc very slightly beyond its borders; more precisely, extend γ slightly beyond p and q in A and push δ out along the radial direction of $TD|_\delta$ i.e. the direction orthogonal to $T\delta$. Now consider the disc bundle $DE \cong D^2 \times D^1$, which is the sub-bundle of the normal bundle of (the extended version of) D determined by the section s, where D coincides with the zero section. The boundary of DE is a 2-sphere, with $\partial(DE) \cap f$ a neighbourhood of γ, that we denote by S. The Whitney move pushes the strip S across DE. The outcome has S replaced by two parallel copies of the Whitney disc D together with a strip whose core is parallel to δ. This is an isotopy of the surface f (if $f \neq g$), and a regular homotopy of $f \cup g$. The latter fact holds since we have described a homotopy through local embeddings. Note that we used a framed and embedded Whitney disc with interior in the complement of $f \cup g$, and the two intersection points p and q were removed, as desired.

In the case that D is framed, but not embedded, or the interior intersects $f \cup g$, the Whitney move, now called a *(framed) immersed Whitney move* still uses the same strip S in a neighbourhood of δ, and two copies of D obtained using s and $-s$, where s is a section of the normal bundle. The resulting move is a regular homotopy of f, and not an isotopy, even if $f \neq g$. We state this fact, but prove it in Section 15.3.

PROPOSITION 11.1. *A (framed) immersed Whitney move is a regular homotopy.*

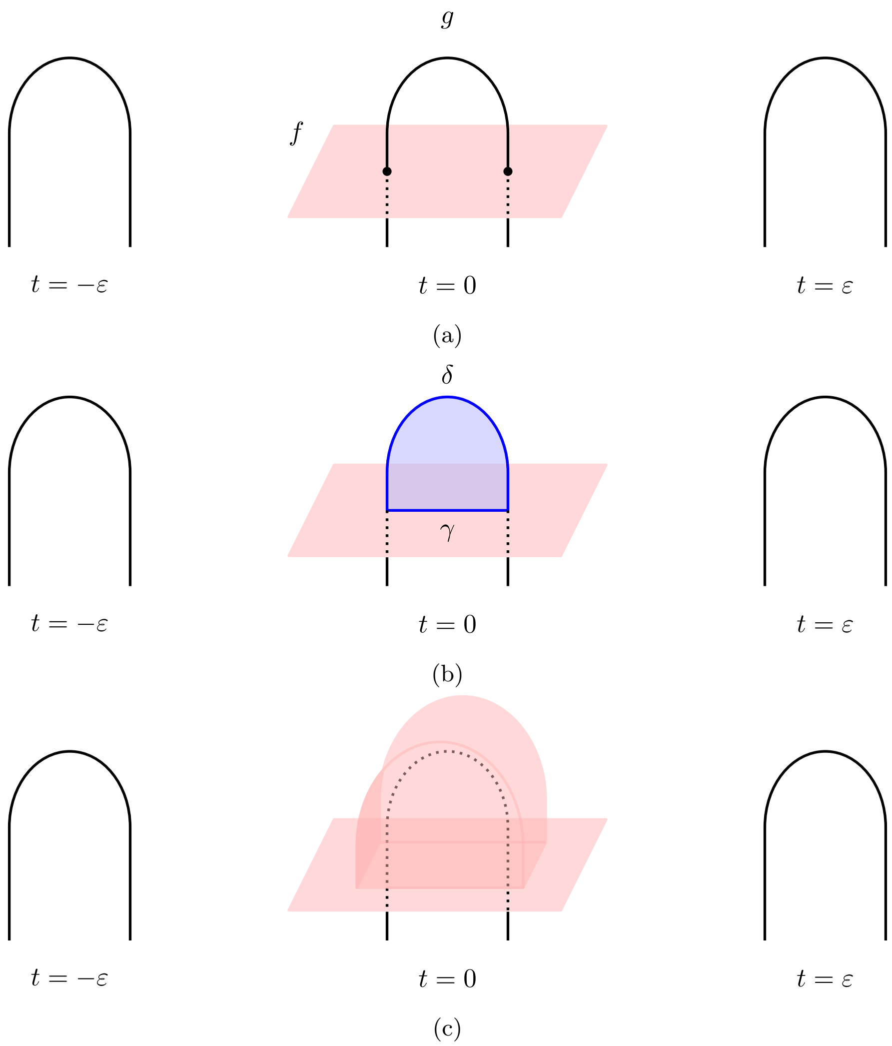


Figure 11.1. A model Whitney move. (a) A region in a smooth 4-manifold M is shown modelled on $\mathbb{R}^3 \times \mathbb{R}$, where the last coordinate t is interpreted as time. The central image shows a small region in an immersed surface f (red). A small region on the immersed surface g (black) is traced out by the black curves as we move backwards and forwards in time.
(b) The two points of intersection between f and g from (a) are shown to be paired by an embedded Whitney disc (blue) with interior in the complement of f and g. The boundary of the disc is given by the union of arcs $\gamma \cup \delta$.
(c) The Whitney move on f along the Whitney disc has removed the two points of intersection.

In particular, the intersection points p and q are removed by an immersed Whitney move, but four new self-intersection points of f are created for each self-intersection point of D, and two new intersections of $f \cup g$ are created for each intersection of the interior of D with $f \cup g$.

In more generality, if D intersects a surface Σ, where Σ may equal f or g, but need not, then two intersection points of f with Σ are created for each intersection point of D with Σ.

11.2.2. Finger moves. As we saw in Chapter 1, a *finger move* is a regular homotopy that adds two intersection points between two immersed surfaces f and g in a smooth 4-manifold M, where possibly $f = g$. It is supported in a neighbourhood of an arc.

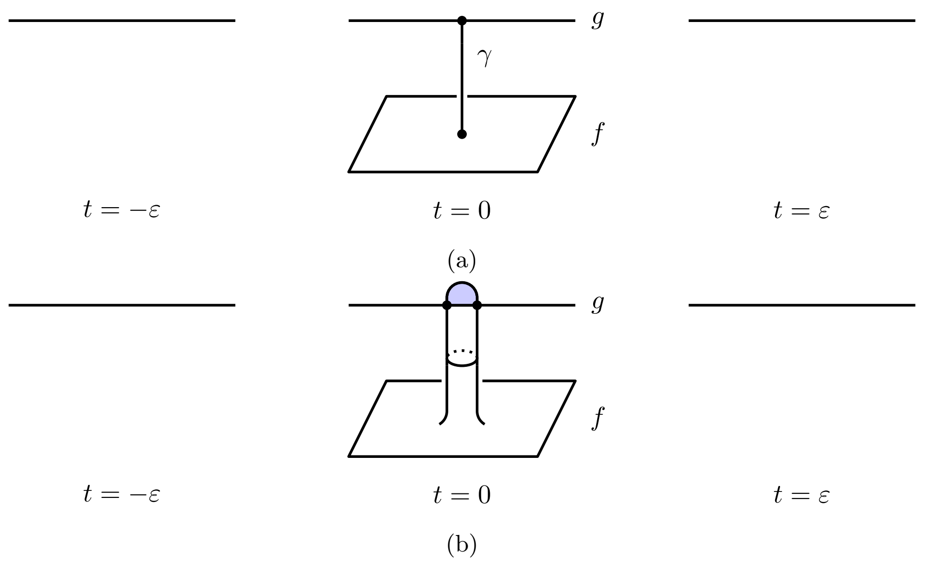


Figure 11.2. A model finger move along the arc γ, shown (a) before and (b) after the move. As usual, the ambient space is shown as $\mathbb{R}^3$ slices moving through time (see the caption to Figure 11.1).

Let $\gamma \cong [0,1]$ be an embedded path in M with an endpoint on f and an endpoint on g, away from any double points, and otherwise disjoint from $f \cup g$. We describe the finger move of f on g along γ. Thicken γ to $\gamma \times D^2$ such that $\{0\} \times D^2 \subseteq f$ is a neighbourhood of one end of γ. Extend γ slightly beyond its other endpoint on g, to an embedding of $[0, 5/4]$. Suppose that $(\{1\} \times D^2) \cap \Sigma$ is a single arc. The finger move pushes $\{0\} \times D^2 \subseteq f$ across $[0, 5/4] \times D^2$, replacing $\{0\} \times D^2$ with the rest of the boundary $([0, 5/4] \times S^1) \cup (\{5/4\} \times D^2)$.

The finger move adds two intersection points between f and g, which are paired by a framed, embedded Whitney disc with interior in the complement of $f \cup g$, should it be required. Note that if $f \neq g$, the finger move is an isotopy of f, and a regular homotopy of $f \cup g$, since we have described a homotopy through local embeddings.

11.3. Intersection and self-intersection numbers

Let M be a smooth 4-manifold. Assume for a moment that M is compact and based and let π denote $\pi_1(M)$ based at the basepoint. The *equivariant intersection form* λ on M is the pairing

$$\begin{array}{rcl} \lambda\colon H_2(M;\mathbb{Z}\pi) \times H_2(M,\partial M;\mathbb{Z}\pi^w) & \to & \mathbb{Z}\pi \\ (x,y) & \mapsto & \langle PD^{-1}(y), x\rangle. \end{array}$$

Here $H_2(M;\mathbb{Z}\pi)$ denotes the second homology of M with twisted coefficients, which is isomorphic to the second homology of the universal cover of M i.e. to $\pi_2(M)$.

The homomorphism $w\colon \pi \to \{\pm 1\}$ is the *orientation character*, which by definition satisfies that $w(\alpha) = -1$ if and only if α is orientation reversing. The orientation character makes the group ring $\mathbb{Z}\pi$ into a left $\mathbb{Z}\pi$-module denoted $\mathbb{Z}\pi^w$ with action $g \cdot r := w(g)gr$, for $g \in \pi$ and $r \in \mathbb{Z}\pi^w$, extended linearly. The Poincaré duality map $PD\colon H^2(M;\mathbb{Z}\pi) \to H_2(M, \partial M; \mathbb{Z}\pi^w)$ is an isomorphism and the pairing denoted $\langle \cdot, \cdot \rangle$ is the Kronecker pairing. The cohomology group with $\mathbb{Z}\pi$ coefficients is isomorphic to cohomology with compact supports of the universal cover of M.

We prefer not to restrict to compact manifolds. To avoid getting into the details of cohomology with compact supports, we will use λ from now on to denote a different notion of intersection number, with a more geometric definition. The notion will be applicable to every smooth, connected 4-manifold, including noncompact and nonorientable 4-manifolds with arbitrarily many boundary components, and coincides with the above definition whenever both apply [**Ran02**, Proposition 7.22]. The new definition will be for intersections between smooth, based immersions of discs or spheres that intersect transversely in double points in their interiors, with no triple points. In particular, discs need not have boundaries mapping to the boundary of M and might not represent homology classes: this is another key reason that we use the geometric definition of λ in this book rather than the homological definition.

In the following definition, and in the rest of the book, we occasionally abuse notation by conflating a map f and its image.

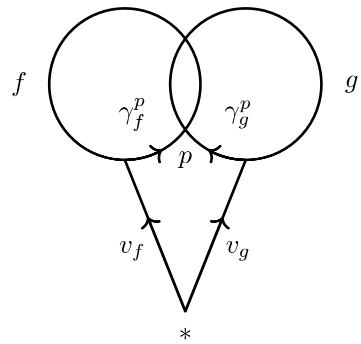


Figure 11.3. Computation of the intersection number $\lambda(f, g)$ for spheres f and g, denoted schematically as circles, in a 4-manifold M. Above $*$ denotes the basepoint of M.

Definition 11.2. Let M be a connected, based, smooth 4-manifold with a fixed local orientation at the basepoint. Let f and g be smoothly immersed, transversely intersecting, based, oriented discs or spheres in M. If applicable assume that f and g have disjointly embedded boundaries. Let v_f and v_g be paths in M joining the basepoint of M to the basepoint of f and g respectively. The paths v_f and v_g are called *whiskers* for f and g respectively. Define the following sum

$$\lambda(f, g) := \sum_{p \in f \pitchfork g} \varepsilon(p)\alpha(p),$$

where

- γ_f^p is a simple path in f from the basepoint of f to p and γ_g^p is a simple path in g from the basepoint of g to p, as in Figure 11.3;
- $\varepsilon(p) \in \{\pm 1\}$ is $+1$ when the local orientation at p induced by the orientations of f and g matches the one obtained by transporting the local orientation at the basepoint of M to p along $v_g\gamma_g^p$, and is -1 otherwise;

- $\alpha(p)$ is the element of $\pi_1(M)$ given by the concatenation $v_f\gamma_f^p(\gamma_g^p)^{-1}v_g^{-1}$.

Since f and g are (immersed) discs or spheres, distinct choices of γ_f^p and γ_g^p are homotopic and thus $\lambda(f,g)$ is a well defined element of $\mathbb{Z}[\pi_1(M)]$.

We also define $\lambda(f,f) := \lambda(f,f^+)$, where f^+ is a push-off of f along a section of the normal bundle transverse to the zero section. If f is an immersed disc with embedded boundary equipped with a specified framing for the normal bundle restricted to the boundary, then f^+ is defined to be the push-off of f along a section restricting to one of the vectors of that framing on ∂f.

Remark 11.3. Note that in the Definition 11.2, we need the simple connectivity of spheres and discs for the intersection number λ to be well defined. In surfaces with nontrivial fundamental group, the choice of path from the basepoint to the double point could change the value of λ, albeit in a controlled manner.

We also note that $\varepsilon(p)$ does not depend on the choice of path γ_g^p when g is an (immersed) disc or sphere, since any two choices of path are homotopic relative the endpoints and thus induce the same local orientation at p. However, $\varepsilon(p)$ might depend on the choice of whisker v_g, for example, if we pre-concatenate v_g with an orientation reversing loop at the basepoint of M.

Define an involution on $\mathbb{Z}[\pi_1(M)]$ by setting

$$h = \sum_{\alpha\in\pi_1(M)} n_\alpha\alpha \mapsto \overline{h} = \sum_{\alpha\in\pi_1(M)} w(\alpha)n_\alpha\alpha^{-1},$$

where $n_\alpha \in \mathbb{Z}$ and $w\colon \pi_1(M) \to \{\pm1\}$ is the orientation character. The next proposition summarises the properties of the intersection number.

Proposition 11.4. *Let M be a connected, based, smooth 4-manifold with a fixed local orientation at the basepoint. Let f and g be smoothly immersed, transversely intersecting, based, oriented discs or spheres in M with whiskers v_f and v_g respectively, and if applicable disjointly embedded boundaries. In particular, f and g only intersect in their interiors. The intersection number $\lambda(f,g)$ has the following properties.*

- *(i) The intersection number $\lambda(f,g)$ is unchanged by regular homotopies in the interiors of f and g. The intersection number is not preserved by a regular homotopy pushing an intersection point off the boundary of an immersed disc, as in Figure 11.4.*
- *(ii) λ is hermitian, that is $\lambda(f,g) = \overline{\lambda(g,f)}$.*
- *(iii) A different choice of whisker v_f' for f results in multiplication of $\lambda(f,g)$ on the left by the element $v_f'v_f^{-1}$ of $\pi_1(M)$. A different choice of whisker v_g' for g results in multiplication of $\lambda(f,g)$ on the right by the element $v_g(v_g')^{-1}$ of $\pi_1(M)$, and changes the sign by $w(v_g(v_g')^{-1})$.*
- *(iv) Changing the orientation of f changes the sign of $\lambda(f,g)$, as does changing the orientation of g. Changing the orientation of both f and g leaves $\lambda(f,g)$ unchanged.*

In the homological version of λ discussed above, (iii) implies that λ is sesquilinear, that is $\lambda(rf,sg) = r\lambda(f,g)\overline{s}$, for $r,s\in\mathbb{Z}[\pi_1(M)]$.

Proof. For (i), we only need to check that the intersection number is preserved under finger moves and Whitney moves along framed, embedded Whitney discs with interiors in the complement of $f\cup g$. By the requirement that the regular homotopies occur in the interiors of f and g, double points are introduced or eliminated in pairs.

Suppose the intersection points p and q between f and g are paired by a framed, immersed Whitney disc, as described in Section 11.2.1. Then there is a path γ_f^{pq} in

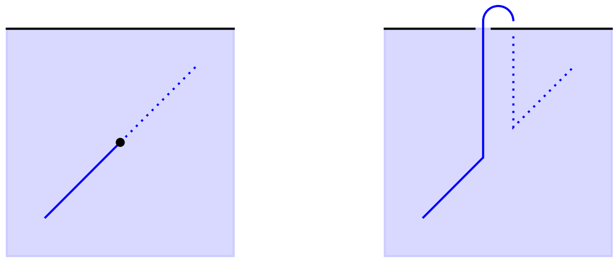


Figure 11.4. Left: A 3-dimensional snapshot of an intersection of an immersed surface f (shown as a blue interval) with an immersed disc g (shaded blue) in an ambient 4-manifold is shown. The interval locally traces out f as we move backwards and forwards in time, as in Figures 11.1 and 11.2.
Right: A finger move on f across the boundary of g removes an intersection point and so changes the intersection number $\lambda(f,g)$.

f from p to q and a path γ_g^{pq} in g from p to q such that $\gamma_f^{pq}(\gamma_g^{pq})^{-1}$ is null-homotopic in M. Moreover, since the Whitney framing exists, $\varepsilon(p)=-\varepsilon(q)$. Let γ_f^p be a path in f from the basepoint of f to p and let γ_g^p be a path in g from the basepoint of g to p. Then the contribution of p and q to $\lambda(f,g)$ is the sum

$$\varepsilon(p)v_f\gamma_f^p(\gamma_g^p)^{-1}{v_g}^{-1}+\varepsilon(q)v_f\gamma_f^p\gamma_f^{pq}(\gamma_g^{pq})^{-1}(\gamma_g^p)^{-1}{v_g}^{-1}$$

which is zero in $\mathbb{Z}[\pi_1(M)]$ since $\varepsilon(p)=-\varepsilon(q)$ and $\gamma_f^{pq}(\gamma_g^{pq})^{-1}$ is null-homotopic in M. Since the Whitney disc is embedded and framed with interior in the complement of $f\cup g$, the Whitney move removes the two intersection points p and q, creates no new intersection points, and preserves the contribution of all other intersection points to $\lambda(f,g)$. It follows that the intersection number $\lambda(f,g)$ is preserved under a Whitney move along framed, embedded Whitney discs with interior in the complement of $f\cup g$. Additionally, this finishes the proof of (i) since a finger move in the interior creates a pair of intersection points paired by an embedded, framed Whitney disc, with interior in the complement of $f\cup g$.

Property (ii) is a direct consequence of the definition

$$\lambda(f,g):=\sum_{p\in f\pitchfork g}\varepsilon(p)\alpha(p),$$

noting that switching the order of f and g replaces every $\alpha(p)$ by $\alpha(p)^{-1}$, while the sign $\varepsilon(p)$ changes precisely when $\alpha(p)$ is orientation reversing.

Property (iii) is also a direct consequence of the definition, since given another whisker v_f' for f, each $\alpha(p)$ changes by multiplication on the left by $v_f'{v_f}^{-1}$, and similarly given another whisker v_g' for g, each $\alpha(p)$ changes by multiplication on the right by $v_g(v_g')^{-1}$. The local orientation is transported along $(v_g'v_g^{-1})v_g\gamma_g=v_g'\gamma_g$ instead of $v_g\gamma_g$, so $\varepsilon(p)$ changes by $w(v_g'v_g^{-1})=w((v_g'v_g^{-1})^{-1})=w(v_g(v_g')^{-1})$.

Property (iv) follows directly from the definition of $\varepsilon(p)$ for a point $p\in f\pitchfork g$. $\square$

Next we define the *self-intersection number* of a smoothly immersed, based, oriented sphere or disc. The definition is analogous to the intersection number λ. However, for the self-intersection number, there is an ambiguity coming from the choice of sheets at each intersection point, as indicated in Proposition 11.4(ii). As a result, the self-intersection number is only well defined as an element of a quotient of $\mathbb{Z}[\pi_1(M)]$, as follows.

Definition 11.5. Let M be a connected, based, smooth 4-manifold with a fixed local orientation at the basepoint. Let f be a smoothly immersed, based, oriented disc or sphere in M with a whisker v, and embedded boundary if applicable. Let $w\colon \pi_1(M) \to \{\pm 1\}$ be the orientation character of M. Define the following sum

$$\mu(f) := \sum_{p \in f \pitchfork f} \varepsilon(p)\alpha(p),$$

where

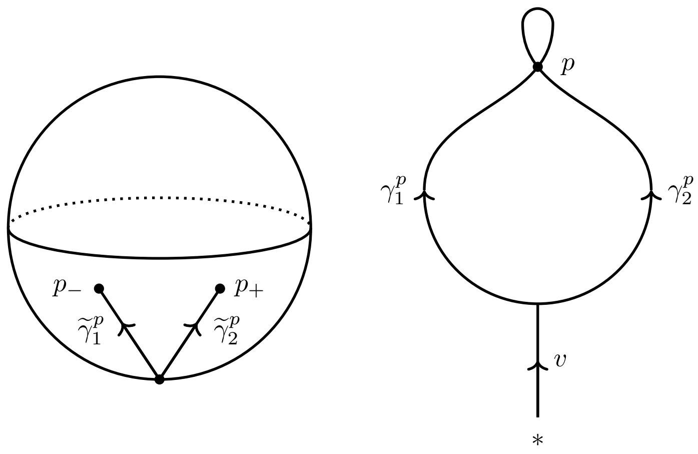


Figure 11.5. Computation of the self-intersection number $\mu(f)$ for an immersed sphere f in a 4-manifold. The point $*$ denotes the basepoint of M.
Left: The points p_+ and p_- in S^2 map to the self-intersection point p of f. Lifts of γ_1^p and γ_2^p to S^2 are shown.
Right: The paths γ_1^p and γ_2^p approach p on two different sheets in the image of f.

- γ_1^p and γ_2^p are simple paths in f from the basepoint to p along two different sheets, as in Figure 11.5;
- $\varepsilon(p) \in \{\pm 1\}$ is $+1$ when the local orientation at p induced by the orientation of f matches the one obtained by transporting the local orientation at the basepoint of M to p along $v\gamma_2^p$, and is -1 otherwise; and
- $\alpha(p)$ is the element of $\pi_1(M)$ given by the concatenation $v\gamma_1^p(\gamma_2^p)^{-1}v^{-1}$.

It follows from the proof of Proposition 11.4(ii) that $\mu(f)$ is a well defined element of the quotient group $\mathbb{Z}[\pi_1(M)]/(a \sim \overline{a})$.

Remark 11.6. As before for λ, we need the simple connectivity of spheres and discs in order for μ to be well defined.

Unlike before, $\varepsilon(p)$ does depend on the choice of the path γ_1^p, even when f is an (immersed) disc or sphere. In particular, the local orientation at p induced by transporting the local orientation at the basepoint of M along $v\gamma_1^p$ and along $v\gamma_2^p$ differ exactly when the loop $\gamma_1^p(\gamma_2^p)^{-1}$ is orientation reversing. Thus in this case the value of $\varepsilon(p)$ depends on the choice of sheet at p. The difference in $\varepsilon(p)$ is encoded within the definition of the involution $\overline{h} = w(h)h^{-1}$, for $h \in \pi_1(M)$. As before, once a sheet is chosen, that is once one of the two preimage points of p is chosen, the value of $\varepsilon(p)$ is well defined, modulo the choice of whisker v.

By virtually the same proof as Proposition 11.4, we have the following properties of the self-intersection number.

PROPOSITION 11.7. *Let M be a connected, based, smooth 4-manifold with a fixed local orientation at the basepoint. Let f be a smoothly immersed, based, oriented disc or sphere in M with a whisker v, and embedded boundary if applicable. The self-intersection number $\mu(f)$ has the following properties.*

(i) *The self-intersection number $\mu(f)$ is unchanged by regular homotopies in the interior of f. The intersection number is not preserved by a regular homotopy pushing an intersection point off the boundary of an immersed disc, as in Figure 11.4.*

(ii) *A different choice of whisker v' for f results in conjugation of $\mu(f)$ by the element $v'v^{-1}$ of $\pi_1(M)$ and multiplication by $w(v'v^{-1})$.*

Moreover, the intersection and self-intersection numbers of an immersed sphere or disc are related by the following helpful formula.

PROPOSITION 11.8. *Let M be a connected, based, smooth 4-manifold with a fixed local orientation at the basepoint. Let f be a smoothly immersed, based, oriented disc or sphere in M. In the case that f is an immersed disc, assume that the boundary is embedded and the normal bundle of the disc restricted to the boundary has a specified framing. Let v be a whisker for f. Then*

$$\lambda(f,f) = \mu(f) + \overline{\mu(f)} + \chi$$

where $\chi \in \mathbb{Z}$ is the Euler number of the normal bundle of f if f is a sphere, or is the twisting number of the framing induced on the boundary by the restriction of the canonical framing of the normal bundle of the immersed disc with respect to the specified framing when f is an immersed disc.

Note above that $\mu(f)$ and $\lambda(f,f)$ do not lie in the same group. However, the formula above still holds since $\mu(f) + \overline{\mu(f)} \in \mathbb{Z}[\pi_1(M)]$ is well defined, i.e. it is independent of the choice of lift of $\mu(f)$ to $\mathbb{Z}[\pi_1(M)]$. Such a lift corresponds to a choice of ordering of the sheets of f at each double point of f.

PROOF. When $\chi = 0$, this is a straightforward consequence of the definitions, observing that each self-intersection point p of f gives rise to a pair of intersection points between f and f^+, where recall that $\lambda(f,f) := \lambda(f,f^+)$ and f^+ is a push-off of f along a section of the normal bundle transverse to the zero section if f is a sphere, and is defined to be the push-off along a section restricting to the specified framing on the normal bundle of the disc restricted to the boundary if f is a disc. The key point is that if one of the new intersection points contributes $\varepsilon(p)\alpha(p)$ to $\lambda(f,f)$, then the other contributes $\overline{\varepsilon(p)\alpha(p)}$.

By the definition of f^+, there are χ intersection points of f and f^+ which do not arise from self-intersection points of f. Such an intersection point r corresponds to the trivial element of $\pi_1(M)$: if γ^r is a path in f joining the basepoint to r, then $\alpha(r)$ may be chosen to be the concatenation $v\gamma^r((\gamma^r)^+)^{-1}(v^+)^{-1}$, where $(\gamma^r)^+$ and v^+ are push-offs of γ^r and v respectively. This concatenation is null-homotopic in M. □

By Proposition 11.8, if f is a immersed sphere with $\lambda(f,f) = 0$ then the Euler number of the normal bundle is even, and so f is homotopic to a map that extends to a framed immersion, via local cusp homotopies. If both $\lambda(f,f)$ and $\mu(f)$ vanish, then the Euler number must already be zero and f extends to a framed immersion. Moreover, we have the following corollary, showing that in many cases $\lambda(f,f) = 0$ implies that $\mu(f) = 0$ for an immersed sphere or disc f.

Corollary 11.9. *Let M be a connected, based, smooth 4-manifold with a fixed local orientation at the basepoint. Suppose that the orientation character vanishes on all the order two elements of $\pi_1(M)$. Let f be a smoothly immersed, based, oriented disc or sphere in M. In case that f is an immersed disc, assume that the boundary is embedded and the normal bundle of the disc restricted to the boundary has a specified framing. Assume that f is framed.*

Then $\lambda(f,f)=0$ implies that $\mu(f)=0$.

Proof. For each equivalence class $\{\alpha,\alpha^{-1}\}\in\pi_1(M)/\sim$, where $\alpha\sim\beta$ if $\alpha=\beta$ or $\alpha^{-1}=\beta$ for any $\alpha,\beta\in\pi_1(M)$, choose a representative $r(\alpha)$. That is, $r(\alpha)$ is either α or α^{-1} (or both) for every $\alpha\in\pi_1(M)$. Write

$$\mu(f)=\sum_{\{\alpha,\alpha^{-1}\}\in\pi_1(M)/\sim} n_\alpha r(\alpha).$$

Then using Proposition 11.8, since f is framed, we have

$$\begin{aligned}0=\lambda(f,f)&=\mu(f)+\overline{\mu(f)}\\ &=\sum_{\{\alpha,\alpha^{-1}\}\in\pi_1(M)/\sim} n_\alpha r(\alpha)+w(\alpha)n_\alpha r(\alpha)^{-1}.\end{aligned}$$

Equating coefficients, and using that if $\alpha=\alpha^{-1}$ then $w(\alpha)=1$ by hypothesis, we see that $n_\alpha=0$ for all $\alpha\in\pi_1(M)$. ☐

On the other hand, if $w(\alpha)=-1$ and $\alpha^2=1$, and f is such that $\mu(f)=\alpha$, then $\lambda(f,f)=0$. So the hypothesis of the corollary cannot be removed.

So far, we have discussed the intersection and self-intersection numbers in general. In the forthcoming proof of the disc embedding theorem, we will primarily use only the following proposition.

Proposition 11.10. *Let M be a connected, smooth 4-manifold. Let f and g be immersed discs or spheres in M intersecting transversely, and with embedded boundaries if applicable.*

1. *The quantity $\lambda(f,g)=0$ for some choice of basepoints for M, f, and g, and some choice of whiskers for f and g, if and only if all the intersection points of f and g can be paired up by framed, immersed Whitney discs in M with disjointly embedded boundaries. These discs may intersect one another, themselves, as well as f and g. Similarly $\lambda(f,g)=1$ for some choice of basepoints for M, f, and g, and some choice of orientations and whiskers for f and g, if and only if all but one intersection point can be paired up in such a manner.*
2. *The quantity $\mu(f)=0$ for some choice of basepoint for M and f, and some choice of whisker for f, if and only if all the self-intersection points of f can be paired up by framed, immersed Whitney discs in M with disjointly embedded boundaries. These discs may intersect one another, themselves, and f.*

Remark 11.11. In the statement of Proposition 11.10, the vanishing of $\lambda(f,g)$ or $\mu(f)$ for some choice of basepoints for M, f, and g, and some choice of whisker for f and g implies the vanishing for all choices of basepoints and whiskers by Propositions 11.4(iii) and 11.7(ii). Thus in these cases it is meaningful to refer to the vanishing of intersection and self-intersection numbers for immersed discs or spheres with no specified choice of basepoints or whiskers. A different choice of whisker or basepoint changes nonzero values of $\lambda(f,g)$ or $\mu(f)$ as dictated by Propositions 11.4(iii) and 11.7(ii). By the Proposition 11.10, from now on, for spheres or discs f and g with no specified choice of whisker or basepoint, when we say $\lambda(f,g)=1$ we mean that all but one of the intersection points may be paired up by Whitney discs in the ambient 4-manifold.

PROOF. Suppose that $\lambda(f,g)=0$ with respect to whiskers v_f and v_g for f and g respectively, corresponding to some chosen basepoints. Then the contributions of the intersection points between f and g to $\lambda(f,g)$ must cancel in pairs. Let p and q be intersection points of f and g such that the contributions of p and q to $\lambda(f,g)$ cancel each other. That is, $\varepsilon(p)=-\varepsilon(q)$ and $\alpha(p)=\alpha(q)$ in $\pi_1(M)$. In other words, $\alpha(p)\alpha(q)^{-1}$ is the trivial element of $\pi_1(M)$. Let γ_f^p be a path in f from the basepoint of f to p and let γ_f^{pq} be a path in f from p to q. Similarly, let γ_g^p be a path in g from the basepoint of g to p and let γ_g^{pq} be a path in g from p to q. We may assume that all these paths are disjointly embedded.Then

$$\begin{aligned}\alpha(p)\alpha(q)^{-1} &= (v_f\gamma_f^p(\gamma_g^p)^{-1}v_g^{-1})(v_g\gamma_g^p\gamma_g^{pq}(\gamma_f^{pq})^{-1}(\gamma_f^p)^{-1}v_f^{-1})\\ &= v_f\gamma_f^p\gamma_f^{pq}(\gamma_g^{pq})^{-1}\gamma_f^{-1}v_f^{-1}\end{aligned}$$

is trivial in $\pi_1(M)$. Thus $\alpha(p)\alpha(q)^{-1}$ is a basepoint-changing conjugation away from the loop $\gamma_f^{pq}(\gamma_g^{pq})^{-1}$, so this loop is also null-homotopic in M. The trace of this null homotopy gives a Whitney disc pairing p and q in M. This gives one direction of the argument. The reverse direction holds since double points paired by a Whitney disc have opposite signs and they have the same element of $\pi_1(M)$ associated with them. A similar argument applies when $\lambda(f,g)=1$.

As we will show in Section 15.2.3, after an isotopy of the Whitney discs and their boundaries, we may arrange for any collection of Whitney circles to be mutually disjoint and embedded. Then by boundary twisting, explained in detail in Section 15.2.2, there is a homotopy of each Whitney disc to a framed Whitney disc. The homotopy is supported in a neighbourhood of one of the boundary arcs. See Remark 15.1.

We have to be a bit more careful when considering the self-intersection number since it takes values in the quotient group $\mathbb{Z}[\pi_1(M)]/(a\sim\overline{a})$. Assume that $\mu(f)=0$ with respect to a whisker v. We now introduce some new notation for the rest of the proof. Given a self-intersection point p of f and an arc γ from the basepoint of f to p, let $\varepsilon_\gamma(p)\in\{\pm1\}$ equal $+1$ when the local orientation at p induced by the orientation of f matches the one obtained by transporting the local orientation at the basepoint of M to p along $v\gamma$, and equal -1 otherwise.

As before, since $\mu(f)=0$, the contributions of the self-intersection points of f to $\mu(f)$ must cancel in pairs. Let p and q be self-intersection points of f such that the contributions of p and q to $\mu(f)$ cancel each other. Then there exist lifts to $\mathbb{Z}[\pi_1(M)]$ of these contributions which cancel each other as elements of $\mathbb{Z}[\pi_1(M)]$. Recall that such a lift to $\mathbb{Z}[\pi_1(M)]$ corresponds precisely to a choice of ordering of the sheets of f at both p and q.

Let γ_i, for $i\in\{1,2,3,4\}$ be disjointly embedded arcs in f such that γ_1 goes from the basepoint of f to p along the second sheet of f at p, γ_2 goes from p to itself leaving on the second sheet of f at p and returning on the first, γ_3 goes from p to q, leaving on the first sheet of f at p and ending on the second sheet of f at q, and γ_4 goes from q to itself, leaving on the second sheet of f at q and returning on the first (see Figure 11.6). Then the contribution of p to $\mu(f)$, lifted to $\mathbb{Z}[\pi_1(M)]$, is

$$\varepsilon(p)\alpha(p)=\varepsilon_{\gamma_1}(p)v\gamma_1\gamma_2\gamma_1^{-1}v^{-1}.$$

The sheet change occurs between γ_2 and γ_1^{-1}. Similarly, the contribution of q to $\mu(f)$, lifted to $\mathbb{Z}[\pi_1(M)]$, is

$$\varepsilon(q)\alpha(q)=\varepsilon_{\gamma_1\gamma_2\gamma_3}(q)v\gamma_1\gamma_2\gamma_3\gamma_4\gamma_3^{-1}\gamma_2^{-1}\gamma_1^{-1}v^{-1},$$

where the sheet change occurs between γ_4 and γ_3^{-1}. Since these lifts cancel in $\mathbb{Z}[\pi_1(M)]$ by the choice of ordering of the sheets, we have $\varepsilon(p)\alpha(p) = -\varepsilon(q)\alpha(q)$. That is,

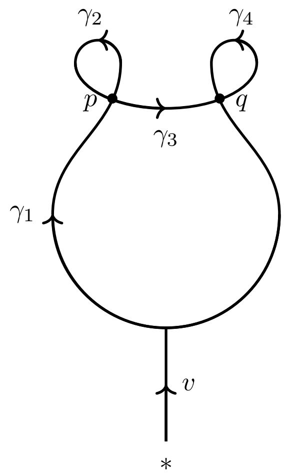


Figure 11.6. Finding Whitney discs pairing self-intersection points of an immersed sphere f when $\mu(f) = 0$.

$$\varepsilon_{\gamma_1}(p) = -\varepsilon_{\gamma_1\gamma_2\gamma_3}(q)$$

and

$$v\gamma_1\gamma_2\gamma_1^{-1}v^{-1} = v\gamma_1\gamma_2\gamma_3\gamma_4\gamma_3^{-1}\gamma_2^{-1}\gamma_1^{-1}v^{-1}$$

in $\pi_1(M)$. In other words,

$$\begin{aligned}(v\gamma_1\gamma_2^{-1}\gamma_1^{-1}v^{-1})&(v\gamma_1\gamma_2\gamma_3\gamma_4\gamma_3^{-1}\gamma_2^{-1}\gamma_1^{-1}v^{-1})\\ &= v\gamma_1\gamma_3\gamma_4\gamma_3^{-1}\gamma_2^{-1}\gamma_1^{-1}v^{-1}\end{aligned}$$

is trivial in $\pi_1(M)$. This is a basepoint-changing conjugation away from the loop $\gamma_3\gamma_4\gamma_3^{-1}\gamma_2^{-1}$, so that loop is null-homotopic in M. The trace of this null homotopy produces a map of a disc in M bounded by the curve $\gamma_3\gamma_4\gamma_3^{-1}\gamma_2^{-1}$. Note that the change of sheets occurs between γ_4 and γ_3^{-1}. We know that $\varepsilon_{\gamma_1}(p) = -\varepsilon_{\gamma_1\gamma_2\gamma_3}(q)$, which implies that $\varepsilon(p) = -\varepsilon(q)$ in the definition of $\mu(f)$. Again, as we shall show in Section 15.2.3, we may arrange for the Whitney circles to be disjointly embedded, and by boundary twisting (Section 15.2.2), for each Whitney disc to be framed (see Remark 15.1). As before, the reverse direction holds since double points paired by a Whitney disc have algebraically cancelling contributions to $\mu(f)$. This completes the proof of Proposition 11.10. □

Given an immersed disc or sphere f in a smooth, connected 4-manifold M, an immersion $g\colon S^2 \looparrowright M$ is said to be a *transverse sphere* or a *dual sphere* for f if f and g intersect transversely and $\lambda(f, g) = 1$. This was called *algebraically transverse* in Chapters 1 and 2. By the above proposition, this means that all but one intersection point between f and g can be paired by Whitney discs in M. If $\lambda(f, g) = 0$, we say that f and g have *algebraically cancelling intersections*. More generally, for a set $\{f_i\}$ of immersed discs or spheres in M, a set of immersed spheres $\{g_i\}$ is said to be a collection of (algebraically) *transverse* or *dual* spheres if for every i, j and some choices of orientations and whiskers, we have that f_i and g_j intersect transversely and $\lambda(f_i, g_j) = \delta_{ij}$. The collection $\{g_i\}$ is said to be a *geometrically transverse* or *geometrically dual* collection of spheres if in addition, $f_i \pitchfork g_j$ is a single point when $i = j$ and is empty otherwise.

11.4. Statement of the disc embedding theorem

Now we state the disc embedding theorem (see [**FQ90**, Theorem 5.1A; **PRT20**]). Recall that since D^2 is contractible, the normal bundle of every immersed disc in a 4-manifold is trivial. For an immersed disc, a choice of orientation of the fibres of the normal bundle determines a framing of the normal bundle, inducing a framing of the normal bundle restricted to the boundary. We say that two immersed discs $f, \overline{f} \colon (D^2, S^1) \looparrowright (M, \partial M)$ have *the same framed boundary* if $f(S^1) = \overline{f}(S^1) \subseteq \partial M$ and there are choices of orientations of the fibres of the normal bundles of f and $\overline{f}$ such that the induced framings on the boundaries are homotopic.

Disc embedding theorem. *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 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.

In other words, within a smooth 4-manifold M with good fundamental group, we can replace a collection of immersed discs $\{f_i\}$, equipped with a collection of transverse spheres $\{g_i\}$ with vanishing intersection and self-intersection numbers, by pairwise disjoint, flat, embedded discs $\{\overline{f}_i\}$ equipped with a collection of geometrically transverse spheres $\{\overline{g}_i\}$, such that for every i, f_i and $\overline{f}_i$ have the same framed boundary.

We do not require M to be compact, since we use the geometric definition of the intersection number λ. In the upcoming proof, we will only use intersection number information to find Whitney discs, as described by Proposition 11.10.

We will define good groups in the next chapter (Definition 12.12), once we have the requisite terminology.

CHAPTER 12

Gropes, towers, and skyscrapers

Mark Powell and Arunima Ray

We define the various iterated objects we will encounter in the proof of the disc embedding theorem, which are called *gropes*, *towers*, and *skyscrapers*.

Roughly, a tower is a sequence of (capped) gropes stacked on top of one another, and a skyscraper is the endpoint compactification of an infinite tower. The construction of skyscrapers, given the data of the disc embedding theorem, via gropes and towers, is the aim of Part II. These objects will be defined as manifolds with corners, so first we briefly recall this notion.

Definition 12.1. Let $\mathbb{R}_{\geq} := \{x \in \mathbb{R} \mid x \geq 0\}$. A (*smooth*) *n-dimensional manifold with corners* M is a Hausdorff, second countable space with a smooth atlas of charts of the following form. To each $x \in M$ is associated an open set $U_x \ni x$ and a chart $\phi_x \colon U_x \xrightarrow{\cong} (\mathbb{R}_{\geq})^k \times \mathbb{R}^{n-k}$ for some k with $0 \leq k \leq n$ and $\phi_x(x) = (0,0)$. Note that the topological manifold boundary of M, which we denote by ∂M, comprises all the points with $k > 0$.

In this book, we shall only consider the cases $n = 4$ and $k \in \{0,1,2\}$, and we call the union of the points with $k = 2$ the *corners* of M. In Definition 12.1, observe that if k is always zero, then M is a smooth manifold with empty boundary. If k is either 0 or 1, then M is a smooth manifold with boundary, where the points with $k = 0$ are precisely the interior points, and the points with $k = 1$ are precisely the boundary points.

Manifolds with corners with $k \in \{0,1,2\}$ arise naturally as products of smooth manifolds with boundary. The canonical example is the square $D^1 \times D^1$ under the product smooth structure. Observe that the square and the disc D^2 are homeomorphic but not diffeomorphic, since one has corners and the other does not. Similarly, the 2-handle $D^2 \times D^2$ naturally has the structure of a manifold with corners, with the decomposition $\partial(D^2 \times D^2) = S^1 \times D^2 \cup_{S^1 \times S^1} D^2 \times S^1$ exhibiting the topological manifold boundary as the union of two smooth boundary regions each diffeomorphic to $S^1 \times D^2$, joined along the corner points $S^1 \times S^1$. As before, note that $D^2 \times D^2$ and D^4 are homeomorphic but not diffeomorphic.

Remark 12.2. Given a manifold with corners, one can obtain a smooth manifold with boundary (that is $k \in \{0,1\}$) by *smoothing corners*. This can be done in an essentially unique way (see [**Wal16**, Proposition 2.6.2] for the case $k \in \{0,1,2\}$). Conversely, given a chosen codimension one submanifold V of the boundary of a smooth manifold with boundary, we may also introduce corners so that V becomes the corner points, in an essentially unique way (see [**Wal16**, Lemma 2.6.3] for the case $k \in \{0,1,2\}$).

When we glue together two manifolds with corners N_1 and N_2, along a common connected component of $\mathrm{cl}(\partial N_1 \setminus \{\text{corners}\})$ and $\mathrm{cl}(\partial N_2 \setminus \{\text{corners}\})$, we identify the corner points ($k = 2$) in such a way that they naturally become smooth boundary points ($k = 1$), as suggested by the symbol $\perp$. See [**Wal16**, Section 2.7] for more details.

We use the terminology of manifolds with corners to give technically correct definitions but we will not emphasise them too much in the sequel. Ultimately, we will lose the vast majority of the smooth information.

12.1. Gropes and towers

Now we begin defining gropes and towers. First we give a somewhat modified definition of the *plumbing* operation from the beginning of Chapter 11. To plumb together two copies of $D^2 \times D^2$, we take small closed disc neighbourhoods D and E in the interior of each copy of $D^2 \times \{0\}$. Choose parametrisations $i_D \colon D^2 \to D$ and $i_E \colon D^2 \to E$ isotopic to $\mathrm{Id} \colon D^2 \to D^2$. Then identify $D \times D^2$ and $E \times D^2$ by setting $(i_D(x), y) \in D \times D^2$ to be equal to $(i_E(y), x) \in E \times D^2$ for a positive plumbing, or equal to $(-i_E(y), -x)$ for a negative plumbing. The resulting identification space is said to be the *result of plumbing* together the two copies of $D^2 \times D^2$. Note that the copies of the original $D^2 \times \{0\}$ intersect in exactly one point in the result, corresponding to the origins $(0, 0)$ of each parametrisation. Finally, smooth corners in the plumbing regions, so that corners only occur where they did in the original copies of $D^2 \times D^2$ prior to plumbing. The same construction can be used to define a self-plumbing of a single copy of $D^2 \times D^2$ in which case D and E are required to be disjoint closed disc neighbourhoods in the interior of $D^2 \times \{0\}$.

In the upcoming discussion, we start with *blocks*, take unions of these to form *stages*, and then stack stages on top of each other to form *generalised towers*. Generalised towers specialise to give *gropes*, *capped gropes*, *towers*, and *capped towers*.

DEFINITION 12.3. A *block* is a quadruple $\mathcal{N} = (N, \Sigma, \phi, \psi)$, where

(i) N is a compact, connected, oriented, smooth 4-manifold with corners with ∂N connected;
(ii) $\Sigma \looparrowright N$ is a properly immersed, oriented surface, called the *spine* of $\mathcal{N}$, with a single boundary component;
(iii) $\phi \colon S^1 \times D^2 \hookrightarrow \partial N$ is a smooth embedding of a solid torus, called the *attaching region* of $\mathcal{N}$, such that $\phi(S^1 \times 0) = \partial\Sigma$;
(iv) $\psi \colon \coprod S^1 \times D^2 \hookrightarrow \partial N$ is a smooth embedding of a (possibly empty) mutually disjoint union of solid tori, called the *tip region(s)* of $\mathcal{N}$, such that $\operatorname{Im}\psi$ is disjoint from $\operatorname{Im}\phi$; and
(v) $\phi(S^1 \times S^1)$ and $\psi\big(\coprod S^1 \times S^1\big)$ form the corners of N.

Often we will abuse notation and refer to the image of ϕ, rather than the map ϕ, as the attaching region of $\mathcal{N}$, and denote it by $\partial_-\mathcal{N}$. Similarly, we will sometimes refer to the image of ψ as the tip region(s) of $\mathcal{N}$. The core of the attaching region is sometimes called the *attaching circle* and the core of a component of the tip region is sometimes called a *tip circle*.

We will use three types of blocks, corresponding to the following standard models.

- A *standard surface block* is a thickened (connected) surface with a single boundary component. We define the elements of the quadruple (N, Σ, ϕ, ψ). Start with a connected, oriented surface Σ_g of genus $g \geq 1$ and a single boundary component embedded in $D^2 \times [0,1] \subseteq \mathbb{R}^3$, with boundary equal to $S^1 \times \{1/2\}$, and thicken to a properly embedded 3-manifold $\Sigma_g \times [0,1]$ in $D^2 \times [0,1]$ with $\partial\Sigma_g \times [0,1] \subseteq S^1 \times [0,1]$. Then take the product of the model with $[0,1]$ to obtain the 4-manifold $\Sigma_g \times [0,1] \times [0,1]$ in $D^2 \times [0,1] \times [0,1]$. The manifold N is the thickened surface $\Sigma_g \times D^2 \hookrightarrow D^2 \times D^2$ obtained by smoothing the corners of $[0,1] \times [0,1]$ to obtain D^2.

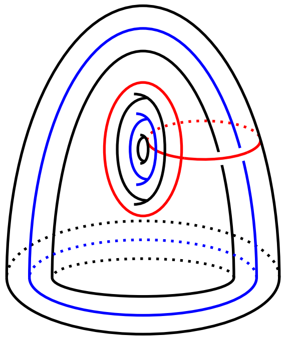


Figure 12.1. A standard surface block with genus one. The spine is shown in blue. The cores of the tip regions, that is the tip circles, are shown in red. The fourth dimension is suppressed.

The core $\Sigma_g \times \{0\}$ is the spine Σ. The boundary S^1 of Σ_g, thickened to a solid torus $S^1 \times D^2$, determines the attaching region ϕ. We thickened in two steps, and the tangent vectors to these two thickenings give a framing and thus an identification of $\partial\Sigma_g \times D^2$ with $S^1 \times D^2$. This determines the map ϕ. Choose a symplectic basis of simple closed curves on Σ_g for $H_1(\Sigma_g;\mathbb{Z})$, and push off to the boundary of $\Sigma_g \times [0,1] \subseteq D^2 \times [0,1]$. For each pair, push one of the curves to $\Sigma_g \times \{0\}$ and the other to $\Sigma_g \times \{1\}$. These are the cores of the tip region. To thicken these cores, and thereby frame the tip region, use one vector that is tangent to $\Sigma_g \times \{j\}$, where j is 0 or 1 as appropriate, and use another vector in the second $[0,1]$ direction of $\Sigma_g \times [0,1] \times [0,1]$. This determines the map $\psi\colon \coprod S^1 \times D^2 \hookrightarrow \partial N$ for the tip regions.

- A *standard disc block* is a self-plumbed thickened disc, with the algebraic count of self-intersections equal to zero. We begin with the thickened disc $D^2 \times D^2$, where the core $D^2 \times \{0\}$ is the spine, and the attaching region ϕ is the inclusion $S^1 \times D^2 \subseteq \partial(D^2 \times D^2)$. The pair (N,Σ) is the image of $(D^2 \times D^2, D^2 \times \{0\})$ after repeated self-plumbings, with algebraically cancelling signs. We require that there are nonzero self-plumbings. The attaching region is preserved by this process. To determine the tip regions we refer to a standard model in Figure 12.2. A pair of algebraically cancelling plumbings is shown. For more plumbings, stack the models on top of one another. Two tip circles are shown in the central picture of the movie. They are by definition *accessory circles*. For the thickening of each of these curves, take one vector that is tangent to the surface shown in the figure as the boundary of N restricted to that time slice, and the other vector in the time direction. This determines (up to isotopy) the map $\psi\colon \coprod S^1 \times D^2 \hookrightarrow \partial N$ for the tip regions. Fix one such map to complete the standard disc block model.
- A *standard cap block* is also a self-plumbed thickened disc, with at least one self-plumbing, but with no restrictions on the signs of the self-plumbings. The attaching region ϕ and the spine Σ is again the image of the inclusion $S^1 \times D^2 \subseteq \partial(D^2 \times D^2)$, and the core $D^2 \times \{0\}$ after repeated self-plumbings. The tip region is empty and therefore there is no need to draw a picture of the standard model. Note that the framing of $S^1 \times D^2$ as the attaching region coincides with the framing induced by the canonical framing of $D^2 \times D^2$ restricted to $S^1 \times D^2$ if and only if the algebraic count of double points vanishes.

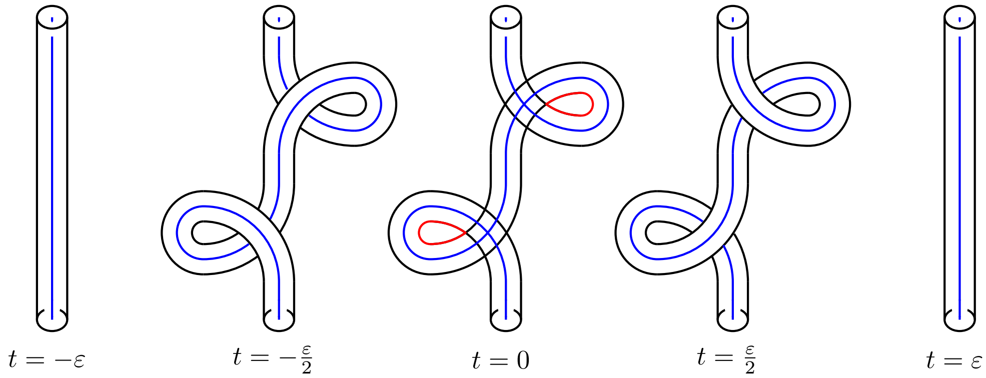


Figure 12.2. Tip regions for a standard disc block. Each panel is a time slice as indicated. The central blue lines in each panel trace out an immersed disc as we move backwards and forwards in time. Black indicates the thickening of the disc in each time slice. The red curves in the central time slice are the accessory circles.

A surface block with $g = 0$ or a disc or cap block with no self-plumbings is the 2-handle $D^2 \times D^2$ with the standard framing on its attaching region and empty tip region, and thus we specifically exclude those cases from our definition.

A block $\mathcal{N} = (N, \Sigma, \phi, \psi)$ is said to be a surface (or disc, or cap) block if there is a diffeomorphism of pairs Θ from the standard surface block (or disc block, or cap block, respectively) to (N, Σ) such that $\phi = \Theta \circ \phi' \colon S^1 \times D^2 \to \partial N$ and $\psi = \Theta \circ \psi' \colon \coprod S^1 \times D^2 \to \partial N$, where ϕ' and ψ' are the attaching and tip regions for the standard model respectively.

Definition 12.4. A *stage* is a nonempty union of surface, disc, and cap blocks $\mathcal{N}_i = (N_i, \Sigma_i, \phi_i, \psi_i)$, denoted $\mathcal{G} = (G := \bigcup N_i, \bigcup \Sigma_i, \Phi := \coprod \phi_i, \Psi)$, such that each surface block and each disc block is disjoint from all other blocks in the stage. Cap blocks are allowed to intersect one another in further arbitrary plumbings. As before $\bigcup \Sigma_i$ is called the *spine* of $\mathcal{G}$ and the smooth embedding $\Phi \colon \coprod S^1 \times D^2 \to \partial G$ arising as the union $\coprod \phi_i$ is called the *attaching region* of $\mathcal{G}$. The *tip region* of a stage is an embedding $\Psi \colon \coprod S^1 \times D^2 \to \partial G$ arising as a union of a *subset* of the maps ψ_i.

Occasionally we will refer to the image of Φ as the attaching region of $\mathcal{G}$, and denote it by $\partial_- \mathcal{G}$. We also refer to the image of Ψ as the tip region(s) of $\mathcal{G}$.

A stage is said to be *homogeneous* if all the blocks are of the same type, that is either all surface blocks, all disc blocks, or all cap blocks. A homogeneous stage is then called either a surface stage, a disc stage, or a cap stage, respectively. Note that a cap stage has empty tip region.

A stage is called *full* if Ψ is the union of all of the ψ_i. A stage that is both homogeneous and full is called *symmetric*.

Finally we stack stages on top of one another to form generalised towers.

Definition 12.5. A *generalised tower* $\mathcal{T}$ of height h is a compact, oriented 4-manifold with corners, with a decomposition

$$G_0 \cup G_1 \cup G_2 \cup \cdots \cup G_h,$$

where $\mathcal{G}_i = (G_i, \Sigma_i, \Phi_i, \Psi_i)$ is a stage for each $i \geq 1$ and $G_0 \subseteq \partial \mathcal{T}$ is a mutually disjoint union $\coprod S^1 \times D^2$ of solid tori, called the *attaching region* of $\mathcal{T}$. We require that

(i) $G_i \cap G_j = \emptyset$ for $|i-j| > 1$;
(ii) G_0 is the attaching region of $\mathcal{G}_1$, that is G_0 is the image of Φ_1 and inherits this framing;
(iii) $\Phi_{i+1} = \Psi_i$ for $i \geq 1$, perhaps after permutation of the connected components of $\coprod S^1 \times D^2$; that is the attaching region of $\mathcal{G}_{i+1}$ and the tip region of $\mathcal{G}_i$ coincide; and
(iv) $\Phi_1(S^1 \times S^1)$ and $\Psi_h\big(\coprod S^1 \times S^1\big)$ form the corners of $\mathcal{T}$.

Often we will denote the attaching region G_0 as $\partial_-\mathcal{T}$. The core of the attaching region G_0 is sometimes called the *attaching circle*. The tip region of $\mathcal{T}$ is by definition the tip region of $\mathcal{G}_h$. The core of a component of the tip region of $\mathcal{T}$ is sometimes called a *tip circle* of $\mathcal{T}$.

A generalised tower is said to be *homogeneous* if each constituent stage is homogeneous, *full* if each constituent stage is full, and *symmetric* if each constituent stage is symmetric. In a homogeneous generalised tower, note that only the last stage $\mathcal{G}_h$ may be a cap stage, since the tip region of a cap block and therefore of a cap stage is empty.

The *spine* of a generalised tower is the 2-complex obtained as the union of the spines $\{\Sigma_i\}$ of the constituent stages together with annuli that join the boundary of Σ_{i+1} to Σ_i, for each i. These annuli are specified in the standard model of a generalised tower, built using standard blocks. Note that the attaching circle is contained in the spine.

Next we restrict to the special types of generalised towers that we will wish to construct in the upcoming proof. In the course of their construction, other generalised towers will arise, so it will be useful to have defined the generalised objects of Definition 12.5.

Definition 12.6.

(1) A *grope* $G(h)$ of height h is a symmetric generalised tower of height h with only surface stages. The spine of a grope is sometimes called a 2-dimensional grope by other authors. See Figure 12.3 and Remark 12.7, compare with Section 5.1.2.

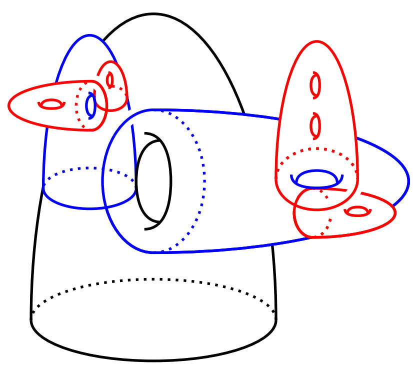


Figure 12.3. Schematic picture of the 2-dimensional spine of a height three grope.

(2) A *capped grope* $G^c(h)$ of height h is a symmetric generalised tower of height $h+1$ where the stages $G_1, \ldots, G_h$ are surface stages and G_{h+1} is a cap stage. The union of the constituent surface stages is called the *body* of the capped grope. See Figure 12.4.

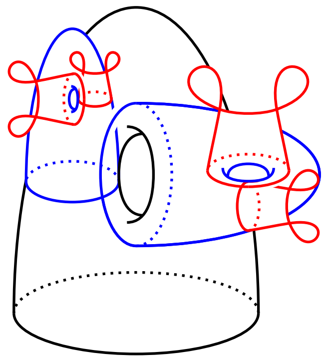


Figure 12.4. Schematic picture of the 2-dimensional spine of a height two capped grope.

(3) A *tower* $\mathcal{T}_n$ with n storeys is a symmetric generalised tower $G_0 \cup G_1 \cup \cdots \cup G_h$, where each $\mathcal{G}_i = (G_i, \Sigma_i, \Phi_i, \Psi_i)$ for $i \geq 1$ is either a surface or disc stage, and such that $\mathcal{G}_h$ is a disc stage. Let $0 = i_0 < i_1 < i_2 < \cdots < i_n = h$ be such that the $\mathcal{G}_{i_j}$, for $j = 1, \ldots, n$, are exactly the disc stages within $\mathcal{T}_n$.
   (a) For $j = 0, \ldots, n-1$, the union $G_{i_j+1} \cup G_{i_j+2} \cup \cdots \cup G_{i_{j+1}}$ is called the $(j+1)$th storey of $\mathcal{T}_n$.
   (b) Note that each storey of $\mathcal{T}_n$ is a capped grope. In particular, the jth storey of $\mathcal{T}_n$ has $i_j - i_{j-1} - 1$ surface stages.
   (c) By definition, the disc stage $\mathcal{G}_{i_j}$ comprises the *jth storey grope caps*. The nth storey grope caps are sometimes referred to as the *top storey grope caps*.

   See Figure 12.5.

(4) A *capped tower* $\mathcal{T}_n^c$ with n storeys is a symmetric generalised tower $G_0 \cup G_1 \cup \cdots \cup G_{h+1}$ where $G_0 \cup G_1 \cup \cdots \cup G_h$ is an n-storey tower and $\mathcal{G}_{h+1}$ is a cap stage.

The boundary $\partial\mathcal{T}$ of a generalised tower contains a submanifold $\partial_-\mathcal{T}$, namely the attaching region. We frequently think of towers as pairs $(\mathcal{T}, \partial_-\mathcal{T})$, since higher towers are intended to be successively better approximations of the pair $(D^2 \times D^2, S^1 \times D^2)$. The tip region of $\mathcal{T}$ is also a submanifold of $\partial\mathcal{T}$, and is denoted $\overline{\partial}\mathcal{T}$. We also have the *vertical boundary* of $\mathcal{T}$, denoted $\partial^+\mathcal{T}$, which is defined to be

$$\partial^+\mathcal{T} := \overline{\partial\mathcal{T} \smallsetminus (\partial_-\mathcal{T} \cup \overline{\partial}\mathcal{T})}.$$

In the upcoming chapters, our goal will be to build the special instances of the generalised towers in the previous definition. However, in order to construct them, as mentioned earlier, along the way we will often use intermediate generalised towers in our proofs, and the intermediate towers will frequently be neither homogeneous nor full. These various generalised towers will be found embedded within the given ambient 4-manifold. However, once we have built a 1-storey capped tower with at least four surface stages, starting in Chapter 18 we will forget the given ambient manifold and work solely within that tower. More precisely, in Chapter 18 we will continue to find embedded generalised towers where the ambient manifold is itself a 1-storey capped tower.

We will often describe operations on gropes and towers within an ambient space in terms of operations on the spines. These are always to be taken to be operations on the 4-dimensional objects, however the operations are usually most easily described

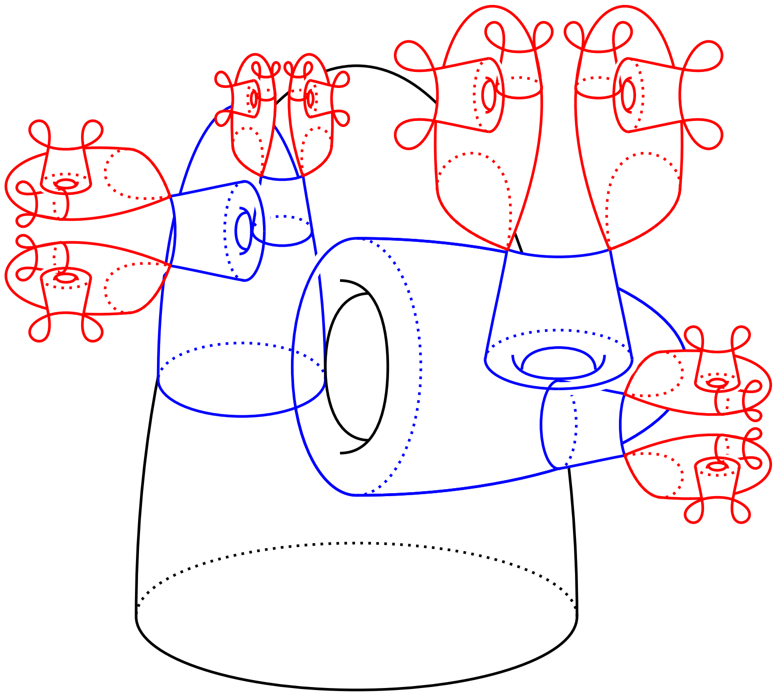


Figure 12.5. Schematic picture of the 2-dimensional spine of a 2-storey tower.

in terms of the spines. The operation can be performed by creating a new spine, and then thickening, by choosing an extension of the spine to an embedding of the associated generalised tower in the ambient space.

Remark 12.7. Gropes often appear in the literature as 2-complexes, with their 4-dimensional thickenings then either called a 'framed grope', or worse, a Grope. This often leads to confusion between the 2- and 4-dimensional objects. In this book, there are no 2-dimensional gropes and so we hope there should be less confusion. In addition, the definition of a grope often given in the literature is of a complex "built" from attaching surfaces or discs. The decomposition into stages is not remembered as part of the structure, but the knowledge of this decomposition is frequently assumed later. But technically one needs to prove that the stage to which a subset of a 2-complex belongs, where the 2-complex was constructed in a certain manner, is a well defined concept. (A similar problem could occur when defining CW complexes.) In Definitions 12.3 and 12.6 we remembered the decomposition into stages as part of the data, so this issue does not arise.

Remark 12.8. In [**FQ90**], the disc stages of a tower have their attaching regions as one Whitney circle and one accessory circle, rather than accessory circles only. However with their choice, the Kirby diagrams change (see Chapter 13), and the resulting links associated with towers in Chapters 13 and 14, are not mixed ramified Bing-Whitehead links, but a slight variation. The Ancel-Starbird result [**AS89**] quoted in [**FQ90**, Section 4.7] does not apply to determine how many surface stages are needed for shrinking. One could instead use the results of [**KP14**] to obtain the correct bound, as explained in Chapter 8. We choose to diverge from [**FQ90**] and attach higher storeys of towers to accessory circles only, in order to avoid (unnecessary) extra complications when describing link diagrams.

The towers/gropes that we defined above are called *union-of-discs-like* (*capped*) *gropes/towers*. In case that the first stage of the grope/tower consists of a single surface block, we have a *disc-like* (capped) grope/tower. A *union-of-spheres-like* (*capped*) *grope/tower* is obtained from a union-of-discs-like (capped) grope/tower by attaching a $D^2 \times D^2$ along $S^1 \times D^2 \subseteq \partial(D^2 \times D^2)$ to each component of the attaching region G_0 using its framing, so that the bottom stage consists of thickened closed surfaces. All the higher surface blocks still have precisely one boundary component. The reader is left the straightforward task of creating the rigorous definitions of this parallel notion, as well as of *sphere-like* (capped) gropes/towers.

Remark 12.9. If a grope or a capped grope has height at least one, it is often useful to have the following additional data. The components of the tip region of the first surface stage admit an assignment of $\{+,-\}$ such that components arising from intersecting curves get different signs. All surface and cap blocks in stages 2 and higher then inherit the assignment of $\{+,-\}$, such that the induced labels on the boundary circles agree in the entire grope. This splits the stages 2 and higher into two gropes, the $(+)$-side and the $(-)$-side, given by all surfaces and caps labelled $+$ and $-$, respectively. The same splitting of higher stages into $(+)$- and $(-)$-sides can be done for towers and capped towers as well, by splitting the base grope, and having all higher stages and storeys inherit signs from below. This is shown in Figure 12.6. Note that $+$ and $-$ are just labels: we could have used colours or letters instead, for example.

We will also need *asymmetric* gropes, that is the height of the $(+)$-side and the $(-)$-side can differ. We make this precise next.

Definition 12.10. An *asymmetric* (*capped*) *grope* is a generalised tower $G = G_0 \cup G_1 \cup G_2 \cup \cdots \cup G_h$, where $G_0 \cup G_1$ is a height one symmetric grope, such that $G_2 \cup \cdots \cup G_h$ has a decomposition as $G^+ \cup G^-$, where G^+ and G^- are symmetric (capped) gropes, called the $(+)$-side and $(-)$-side of G respectively. Let G_0^+ and G_0^- denote the attaching region of G^+ and G^-, respectively. We require that the cores of the attaching regions G_0^+ and G_0^- are isotopic in G_1 to dual halves of a symplectic basis of simple closed curves for $H_1(\Sigma_1;\mathbb{Z})$ for the spine Σ_1 of G_1.

If G^+ and G^- are (capped) gropes of height a and b respectively, we say G has height (a,b).

Remark 12.11. As gropes of height $(b, b-1)$ (or of height $(b-1, b)$) can be understood as lying half way between heights b and $b+1$, they are said to have height $b+0.5$. As a clarifying example, a grope of height $(1,0)$ is often denoted by $G(1.5)$ and is said to have height 1.5. A symmetric grope $G(3)$ of height 3 with a sign assignment as described earlier can be viewed as an asymmetric grope of height $(2,2)$ by taking the $(+)$-side and the $(-)$-side to be the union of all surfaces labelled $+$ and $-$, respectively.

We are now finally able to precisely define good groups. In the following definition, a *double point loop* arises as follows. Given an immersed disc in a 4-manifold, choose a path in the domain D^2 between the pre-image points of a double point, that is the endpoints of the path are mapped to the double point by the immersion. The image of the path is said to be a double point loop. Any two choices of such a path on D^2 are homotopic relative to the boundary, so the resulting double point loops are homotopic relative to the double point. An accessory circle of a disc block of a grope is isotopic to a double point loop for the immersed spine D^2.

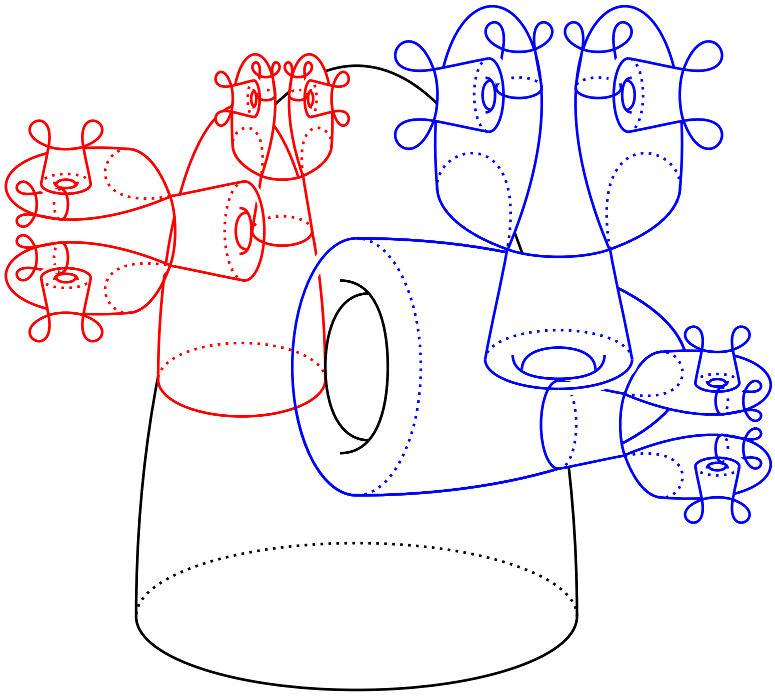


Figure 12.6. Splitting the higher stages of a 2-storey tower into $(+)$- and $(-)$-sides. The $(+)$-side is in blue and the $(-)$-side is in red. Only the 2-dimensional spine is shown.

Definition 12.12. A group Γ is said to be *good* if for every disc-like capped grope $G^c(1.5)$ of height 1.5 with some choice of basepoint, and every group homomorphism $\phi\colon \pi_1(G^c(1.5)) \to \Gamma$, there exists an immersed disc $D^2 \looparrowright G^c(1.5)$, whose framed boundary coincides with the attaching region of $G^c(1.5)$, such that the elements in $\pi_1(G^c(1.5))$ given by its double point loops, considered as fundamental group elements by making some choice of basing path, are mapped to the identity element of Γ by ϕ.

We end this section with two more definitions. From now on, whenever we discuss intersections between gropes/tower stages, or between grope stages and another surface in an ambient 4-manifold, we always talk about intersections with the spine in a union of isolated points. This can be arranged by transversality and means that there is a plumbing of the stages/surfaces – this can be made precise but it would be unnecessarily cumbersome to always describe operations and intersections in these technically correct terms.

Definition 12.13. Let M be a connected, smooth 4-manifold. Let $\mathcal{A}$ be an embedded disc-like or sphere-like generalised tower in M. A *geometrically transverse sphere* for $\mathcal{A}$ is a framed immersed sphere in M intersecting the bottom stage G_1 in a single point, and otherwise disjoint from $\mathcal{A}$.

More generally, let $\mathcal{A}$ be a union-of-discs-like or union-of-spheres-like generalised tower in M, for example the union of a collection of embedded disc-like or sphere-like generalised towers. Let $\{G_1^i\}$ denote the collection of blocks in the bottom stage of $\mathcal{A}$. A collection $\{\Sigma_i\}$ of immersed, framed spheres in M is said to be *geometrically transverse* to $\mathcal{A}$ if $G_1^i \pitchfork \Sigma_j$ is a single point when $i = j$ and is empty otherwise, and if moreover $\{\Sigma_i\}$ has no intersection with any higher stages of $\mathcal{A}$, including caps.

We will also need the following generalisation of geometrically transverse spheres.

Definition 12.14 (Geometrically transverse capped gropes and surfaces). Let M be a connected, smooth 4-manifold. Let $\mathcal{A}$ be an embedded generalised tower in M. A collection $\{T_i^c\}$ of sphere-like (capped) gropes in M is said to be *geometrically transverse* to $\mathcal{A}$ if $A_1^i \pitchfork T_j^c$ is a single point in the bottom stage of T_j^c when $i = j$ and is empty otherwise, where $\{A_1^i\}$ is the collection of blocks in the first stage of $\mathcal{A}$, and moreover $\{T_i^c\}$ has no intersections with higher stages of $\mathcal{A}$. The caps of $\{T_i^c\}$ are disjoint from $\mathcal{A}$ and intersections are allowed within the caps of the collection $\{T_i^c\}$, but the bodies are required to be disjointly embedded, unless otherwise specified.

If each T_i^c is a sphere-like capped grope of height one, then we say that $\{T_i^c\}$ is a collection of *geometrically transverse capped surfaces* for $\mathcal{A}$.

12.2. Infinite towers and skyscrapers

The compact objects defined in the previous section have the following natural extensions.

Definition 12.15. An *infinite generalised tower* $\mathcal{T}_\infty$ is an oriented 4-manifold with corners, with a decomposition

$$\mathcal{T}_\infty = \bigcup_{i=0}^{\infty} G_i$$

where $\mathcal{G}_i = (G_i, \Sigma_i, \Phi_i, \Psi_i)$ is a stage for each $i \geq 1$ and $G_0 \subseteq \partial\mathcal{T}$ is a mutually disjoint union $\coprod S^1 \times D^2$ of solid tori, called the *attaching region* of $\mathcal{T}_\infty$. We require that

(i) $G_i \cap G_j = \emptyset$ for $|i - j| > 1$;
(ii) G_0 is the attaching region of $\mathcal{G}_1$, that is G_0 is the image of Φ_1 and inherits this framing;
(iii) $\Phi_{i+1} = \Psi_i$ for $i \geq 1$, perhaps after permutation of the connected components of $\coprod S^1 \times D^2$; that is the attaching region of $\mathcal{G}_{i+1}$ and the tip region of $\mathcal{G}_i$ coincide; and
(iv) $\Phi_1(S^1 \times S^1)$ forms the corner of $\mathcal{T}_\infty$.

An infinite generalised tower is said to be *homogeneous* if each constituent stage is homogeneous, *full* if each constituent stage is full, and *symmetric* if each constituent stage is symmetric.

The *spine* of an infinite generalised tower is the 2-complex obtained as the union of the spines $\{\Sigma_i\}$ of the constituent stages together with annuli that join the boundary of Σ_{i+1} to Σ_i for each i. These annuli are specified in the standard model of an infinite generalised tower.

The boundary $\partial\mathcal{T}_\infty$ of an infinite generalised tower contains the attaching region $\partial_-\mathcal{T}_\infty$ as a submanifold. We also have the *vertical boundary* of $\mathcal{T}_\infty$, denoted $\partial^+\mathcal{T}_\infty$, which is defined to be

$$\partial^+\mathcal{T}_\infty := \overline{\partial\mathcal{T}_\infty \setminus \partial_-\mathcal{T}_\infty}.$$

We will only use a particular type of infinite generalised tower, which we define next.

Definition 12.16. An *infinite tower* $\mathcal{T}_\infty$ is a symmetric infinite generalised tower $\bigcup_{i=0}^{\infty} G_i$, where each $\mathcal{G}_i = (G_i, \Sigma_i, \Phi_i, \Psi_i)$ for $i \geq 1$ is either a surface or disc stage.

Let $\{i_j\}_{j=0}^{\infty}$ be a strictly increasing sequence with $i_0 = 0$ such that $\mathcal{G}_{i_j}$, for $j \geq 1$, are exactly the disc stages of $\mathcal{T}_\infty$.

(a) For $j \geq 0$, the union $G_{i_j+1} \cup G_{i_j+2} \cup \cdots \cup G_{i_{j+1}}$ is called the $(j+1)$th storey of $\mathcal{T}_\infty$.
(b) Note that each storey of $\mathcal{T}_\infty$ is a capped grope. In particular, the jth storey has $i_j - i_{j-1} - 1$ surface stages.
(c) By definition, the disc stage $\mathcal{G}_{i_j}$ comprises the *jth storey grope caps*.

For an infinite tower $\mathcal{T}_\infty$ and $k \in \mathbb{N}$, we use $\mathcal{T}_\infty^{\leq k}$ to denote the finite tower obtained as a truncation, consisting of the first k storeys of $\mathcal{T}_\infty$.

The object we have defined is a *union-of-discs-like* infinite tower. We can similarly define *disc-like*, *sphere-like*, and *union-of-spheres-like* infinite towers as before. We leave these to the reader.

As defined, an infinite tower $\mathcal{T}_\infty$ is a noncompact space. We now consider its *endpoint compactification*, denoted by $\widehat{\mathcal{T}}_\infty$, which we call an *infinite compactified tower*. The definition of endpoint compactification, given below, is due to Freudenthal [**Fre31**, **Fre42**] (see also [**Pes90**]). In the proof of the disc embedding theorem we will construct embedded infinite compactified towers in the ambient 4-manifold, by squeezing higher storeys of the infinite tower into smaller and smaller balls. The denouement of the proof in Part IV will consist of showing that certain infinite compactified towers are in fact homeomorphic, relative the boundary, to the standard 2-handle.

Definition 12.17 (Endpoint compactification). Let X be a locally compact topological space that is the union of an ascending sequence of compact subsets

$$K_1 \subsetneq K_2 \subsetneq \cdots ,$$

namely an *exhaustion by compact sets* with K_i contained in the topological interior $\operatorname{Int} K_{i+1}$ for each i. The complementary sets $C_i := X \smallsetminus K_i$ form the inverse sequence

$$C_1 \supsetneq C_2 \supsetneq \cdots .$$

Let $E(X)$ be the set of sequences $(U_1, U_2, \dots)$ of nonempty open sets such that

$$U_1 \supsetneq U_2 \supsetneq \cdots$$

and where each U_i is a connected component of C_i. Each element of $E(X)$ is called an *end* of the space X. The *endpoint compactification of* X, denoted $\widehat{X}$, is defined to be $X \cup E(X)$, with the topology generated by open sets of X together with all sets $V \subseteq \widehat{X}$ such that $V \cap X = U_j$ for some j and $V \cap E(X)$ consists of sequences $(V_1, V_2, \dots)$ such that $V_{j+i} \subseteq U_j$ for all $i > 0$.

For example, the real line has two ends, and every compact space has no ends. Informally, the ends of a space correspond to essentially distinct rays to infinity, and the endpoint compactification of a space is obtained by adding a point for each end. We remark that the definition given above differs from that in [**FQ90**, p. 60]. Ours is the more commonly used definition and is better suited to our work with towers. We observe in passing that every topological manifold has an exhaustion by compact sets [**Lee11**, Proposition 4.76] as needed by the definition above.

Remark 12.18. The cardinality of the set of ends of a space and the homeomorphism type of its endpoint compactification do not depend on the choice of exhausting sequence of compact sets. For an infinite tower, when considering the endpoint compactification, we will only ever use the exhausting sequence of compact sets corresponding to finite truncations to towers, and this issue will not arise.

REMARK 12.19. Let $\mathcal{T}_\infty$ be an infinite tower. Certainly $\mathcal{T}_\infty$ is a subset of its endpoint compactification $\widehat{\mathcal{T}}_\infty$, and $\mathcal{T}_\infty$ is a manifold (with corners). However, the space $\widehat{\mathcal{T}}_\infty$ may not be a manifold. The *attaching region* of $\widehat{\mathcal{T}}_\infty$, denoted by $\partial_-\widehat{\mathcal{T}}_\infty$, is defined to be the attaching region of the subset $\mathcal{T}_\infty$. Moreover, note that $\mathcal{T}_\infty^{\leq k} \cap \partial_+\mathcal{T}_\infty = \partial_+\mathcal{T}_\infty^{\leq k}$ for all k. Thus the exhaustion of $\mathcal{T}_\infty$ by compact sets given by the finite truncations induces an exhaustion of $\partial_+\mathcal{T}_\infty$ by compact sets. As a result, there is a canonical bijective correspondence between the ends of $\partial_+\mathcal{T}_\infty$ and the ends of $\mathcal{T}_\infty$. By inspection of the definition, we see that the endpoint compactification of $\partial_+\mathcal{T}_\infty$, called the *vertical boundary* of $\widehat{\mathcal{T}}_\infty$ and denoted by $\partial_+\widehat{\mathcal{T}}_\infty$, is a subset of the endpoint compactification $\widehat{\mathcal{T}}_\infty$ of $\mathcal{T}_\infty$. The vertical boundary of $\widehat{\mathcal{T}}_\infty$ is the union of the vertical boundary of the underlying infinite tower $\mathcal{T}_\infty$ and the endpoints of $\mathcal{T}_\infty$. The union

$$\partial\widehat{\mathcal{T}}_\infty := \partial_-\widehat{\mathcal{T}}_\infty \cup \partial_+\widehat{\mathcal{T}}_\infty$$

is called the *boundary* of $\widehat{\mathcal{T}}_\infty$. The use of the word 'boundary' is somewhat wishful. In particular, it does not refer to the topological boundary. *A priori* we do not know that $\widehat{\mathcal{T}}_\infty$ is a manifold, and therefore we have no clear notion of its boundary.

It is not too hard to argue that if a topological space X is Hausdorff, connected, locally connected, and locally compact, then $\widehat{X}$ is compact Hausdorff. By the Urysohn metrisation theorem, a compact Hausdorff space is metrisable if and only if it is second countable. Since every infinite tower $\mathcal{T}_\infty$ is a manifold, it has a countable basis $\mathcal{B}$ for its topology. By definition, a basis for the topology on $\widehat{\mathcal{T}}_\infty$ is given by the union of $\mathcal{B}$ with the collection of components of complements of the finite towers obtained as truncations of $\mathcal{T}_\infty$. Since the latter collection is countable, we have established the following fact.

PROPOSITION 12.20. *Every infinite compactified tower is metrisable.*

Note that the number of ends of an infinite tower is uncountable, since there are at least two branches at every stage of the tower, but we still obtain a countable basis for the topology. This is not a contradiction, since every open set containing an endpoint contains infinitely many other endpoints. The situation is thus markedly different from simple examples of endpoint compactifications such as the real line or an infinite cylinder. Nonetheless, any two endpoints of an infinite tower can be separated by small enough open sets, corresponding to taking the complement of a tall enough finite truncation. This is another way to see that every infinite compactified tower is Hausdorff.

When an infinite tower $\mathcal{T}_\infty$ is embedded in an ambient 4-manifold M, it is said to *converge* if the closure of $\mathcal{T}_\infty$ in M is equal to its endpoint compactification $\widehat{\mathcal{T}}_\infty$. These were called *convergent infinite towers* in [**FQ90**].

We now define a particular type of infinite compactified tower that will be a key ingredient in our proof of the disc embedding theorem.

DEFINITION 12.21 (Skyscrapers and open skyscrapers). A disc-like infinite tower $\mathcal{S}$ is said to be an *open skyscraper* if it satisfies the following two conditions.

(a) (Replicable) Each storey of $\mathcal{S}$ has at least four surface stages.
(b) (Boundary shrinkable) The series $\sum_{i=1}^{\infty} N_j/2^j$ diverges, where N_j is the number of surface stages in the jth storey of $\mathcal{S}$.

The union of the $2i-1$ and the $2i$ storeys of an open skyscraper is called its ith *level*.

The endpoint compactification of an open skyscraper $\mathcal{S}$ is called a *skyscraper*, denoted by $\widehat{\mathcal{S}}$. Similarly, given a skyscraper $\widehat{\mathcal{S}}$, the corresponding open skyscraper is denoted by $\mathcal{S}$.

For a skyscraper $\widehat{\mathcal{S}}$, we use $\widehat{\mathcal{S}}^{\leq k}$ to denote the finite tower obtained as a truncation, consisting of the first k levels of $\widehat{\mathcal{S}}$, and $\widehat{\mathcal{S}}^{k}$ to denote the finite tower consisting of exactly the kth level.

Remark 12.22. Observe that since a skyscraper is a particular type of infinite compactified tower, by Proposition 12.20 every skyscraper is metrisable.

The motivation behind the replicable condition in the definition above will become clear in the forthcoming chapters, specifically Chapter 18. The boundary shrinkable condition implies, via Theorem 8.1, that a corresponding mixed ramified Bing-Whitehead decomposition shrinks (the correspondence will be explained in Chapter 13), and as a result we infer that the boundary of every skyscraper $\widehat{\mathcal{S}}$ is first of all a manifold, and secondly homeomorphic to S^3. Recall that in order to prove the disc embedding theorem, we will show that every skyscraper is homeomorphic, relative to the attaching region, to the standard 2-handle. The fact that the boundary of $\widehat{\mathcal{S}}$ is homeomorphic to S^3 is therefore highly encouraging since it reveals that at least we start with the correct boundary. We will understand much more about the structure of skyscrapers, their boundaries, and their truncations in the following two chapters.

CHAPTER 13

Picture camp

Duncan McCoy, JungHwan Park, and Arunima Ray

We begin this chapter by introducing *Kirby diagrams* and *Kirby calculus*. For a more detailed account, see [**GS99**]. Then we describe how to build simple Kirby diagrams for the iterated spaces from the previous chapter, namely gropes, capped gropes, towers, and capped towers. An expert in Kirby calculus may choose to skip to the summary in Section 13.7. The title of this chapter is in honour of Freedman's trip to "Picture camp" in his 2013 lecture series.

13.1. Dehn surgery

Let $K \subseteq S^3$ be a knot. Then a *closed* tubular neighbourhood νK of K is diffeomorphic to $S^1 \times D^2$. Given a diffeomorphism $\phi\colon \partial\nu K \to S^1 \times S^1$, the result of *Dehn surgery* on S^3 along K with respect to ϕ is the 3-manifold

$$\overline{S^3 \setminus \nu K} \cup_\phi S^1 \times D^2.$$

A 0-*framed longitude* of K, sometimes called the *Seifert longitude* of K, is a parallel push-off of K that has trivial linking number with K, or equivalently, is null-homologous in the complement of K in S^3. This uniquely determines a parallel push-off of K up to homotopy in $\partial\nu K$. The solid torus being glued in is called the *surgery solid torus*. Let $\phi\colon \partial\nu K \to S^1 \times S^1$ be the map, unique up to isotopy, taking the 0-framed longitude of K to a meridian $\{*\} \times \partial D^2$. The Dehn surgery on S^3 along K with respect to this choice of ϕ is said to be 0-framed. Similarly, one performs 0-framed Dehn surgery on S^3 along a link L as removing a collection of pairwise disjoint tubular neighbourhoods of the components of L from S^3 and then gluing in solid tori so that the 0-framed longitude of each component is identified with a meridian of the glued in surgery solid tori. The resulting 3-manifold is represented by a *surgery diagram* consisting of a diagram of L in S^3 where each component of L is decorated with a zero.

13.2. Kirby diagrams

Fix $n \geq 0$. An n-dimensional k-handle h, for $0 \leq k \leq n$, is a copy of $D^k \times D^{n-k}$. The disc $D^k \times \{0\}$ is the *core* of h and the disc $\{0\} \times D^{n-k}$ is the *cocore*. The boundary ∂h of h naturally decomposes as the union of $\partial_- h := S^{k-1} \times D^{n-k}$, called the *attaching region* of h, and $\partial_+ h := D^k \times S^{n-k-1}$. The boundary of the core, namely $S^{k-1} \times \{0\}$, is called the *attaching sphere* of h and the boundary of the cocore, namely $\{0\} \times S^{n-k-1}$, is called the *belt sphere*. When $k = 2$, the attaching sphere is sometimes called the *attaching circle*. The integer k is called the *index* of h. We restrict ourselves henceforth to $n = 4$. In fact, in this book we will only need to concern ourselves with $k \in \{0, 1, 2\}$. Figure 13.1 shows a 3-dimensional ($n = 3$) 1- and 2-handle.

Note than handles of a fixed dimension are all homeomorphic and, modulo smoothing corners, diffeomorphic. What specifies a handle of a given index is how it is attached. Given a smooth 4-manifold X with nonempty boundary and a

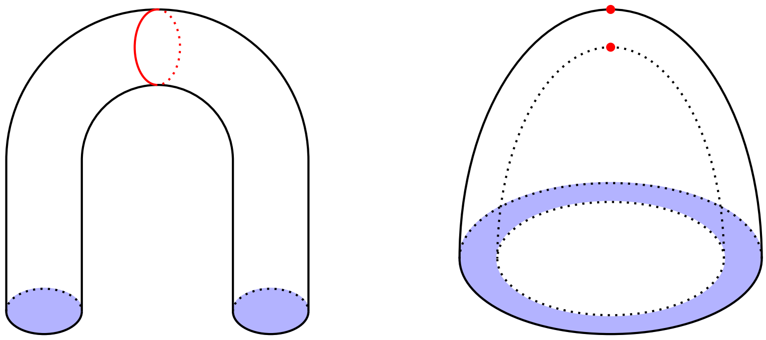


Figure 13.1. A 3-dimensional 1-handle $D^1 \times D^2$ and 2-handle $D^2 \times D^1$. The attaching regions and the belt spheres are drawn in blue and red, respectively.

smooth embedding $\phi\colon S^{k-1} \times D^{4-k} \hookrightarrow \partial X$, we attach a 4-dimensional k-handle to X by forming the identification space

$$X' := \left(X \sqcup D^k \times D^{4-k}\right)/x \sim \phi(x), \forall x \in S^{k-1} \times D^{4-k},$$

where we smooth the corners produced by the identification (see Remark 12.2). By construction, the resulting space X' is also a smooth 4-manifold, possibly with boundary, but with no corners. The diffeomorphism type of X' depends on the isotopy class of ϕ, but is otherwise well defined. The isotopy class of ϕ is determined by an embedding $S^{k-1} \hookrightarrow \partial M$ and a choice of framing for the normal bundle of the image. The choices of framing for the normal bundle are in bijection with elements of $\pi_{k-1}(O(4-k))$.

A *handle decomposition* of a connected, smooth 4-manifold X is an identification of X with a manifold obtained by iteratively attaching handles to the empty space $\emptyset$. In particular, a handle decomposition for any nonempty manifold must have at least one 0-handle, which is the only type of handle with empty attaching region. It follows from Morse theory (e.g. [**Mil63**, **Mil65**]), that any connected, smooth manifold X admits a handle decomposition. Moreover, we can assume that there is a single 0-handle, that handles are attached in increasing order of index, and that handles of the same index are attached in any order or simultaneously, as we wish. In particular, we can assume that the images of the attaching regions of handles with index k lie in the $(k-1)$-skeleton of X, for every k.

For us, this means that any connected, smooth 4-manifold is obtained by attaching 1-, 2-, 3- and 4-handles to a 0-handle D^4 along $\partial D^4 = S^3$. We will soon see how to describe the attaching regions of such handles. The spaces we are interested in can be described using only 1- and 2-handles (in addition to the 0-handle), so we restrict ourselves to those cases. The attaching regions will be described within S^3, considered to be the boundary of a 0-handle. In practice, we will draw pictures in $\mathbb{R}^3$, with the understanding that S^3 is the one-point compactification. Such diagrams in $\mathbb{R}^3$ of attaching regions of handles for a 4-manifold are sometimes referred to as *Kirby diagrams*.

We mention in passing that there is also a theory of *relative handle decompositions*. That is, given a manifold X and $\partial_- X$ a component of the boundary of X, a relative handle decomposition of $(X, \partial_- X)$ is an identification of X with a manifold obtained by iteratively attaching handles to $\partial_- X \times [0,1]$ along $\partial_- X \times \{1\}$.

It is straightforward to see from the definitions that given a smooth 4-manifold X, a component $\partial_- X$ of ∂X, and a relative handle decomposition $\mathcal{H}$ of $(X, \partial_- X)$, there is a corresponding relative handle decomposition $\mathcal{H}'$ of $(X, \partial X \smallsetminus \partial_- X)$, where each

k-handle of $\mathcal{H}$ becomes a $(4-k)$-handle of $\mathcal{H}'$. The decomposition $\mathcal{H}'$ is said to the result of turning $\mathcal{H}$ *upside down*.

13.2.1. Attaching 1-handles. As mentioned previously, we will assume that handle decompositions have a single 0-handle and that handles are attached in increasing order of index. Thus, 1-handles are attached to the boundary S^3 of a single 0-handle D^4. Recall that a 4-dimensional 1-handle is a copy of $D^1 \times D^3$. The attaching region is a copy of $S^0 \times D^3$ and the attaching sphere is a copy of S^0. This means that up to isotopy there are precisely two ways to attach a 1-handle to a 0-handle, corresponding to the two connected components of $O(3)$. One of these results in the nonorientable D^3 bundle over S^1 and the other yields $S^1 \times D^3$. Thus there is an essentially unique way to attach a 1-handle to a 0-handle in order to obtain an orientable manifold. If one attaches n such orientable 1-handles to a 0-handle, then the resulting 4-manifold is the boundary connected sum $\natural^n S^1 \times D^3$, with boundary $\#^n S^1 \times S^2$.

There are two different notations to denote a 1-handle attachment. The first of these is to directly draw the attaching region of a 1-handle as two solid balls in $\mathbb{R}^3$, which are then identified by an orientation-reversing diffeomorphism, for example, by a reflection in a plane equidistant from the two balls, when the 1-handle is attached. In this notation, a loop travelling along the 1-handle appears as an arc which emerges from one of these balls and enters the other.

The alternative notation for 1-handle attachments arises from observing that $S^1 \times D^3$ is obtained as the complement of the standard unknotted disc in D^4 bounded by an unknot in the boundary S^3. In other words, if one takes an embedded disc in S^3 and pushes its interior into D^4, then the complement of an open tubular neighbourhood of this disc is $S^1 \times D^3$. Repeating this process with n pairwise disjoint discs yields $\natural^n S^1 \times D^3$. As a result, we may also indicate 1-handles attached to a 0-handle by drawing the trivial link in $\mathbb{R}^3$ bounding these pushed in discs. Each component of the unlink is decorated by a dot, to distinguish from the notation for 2-handles which we will soon describe. In this notation for a 1-handle, a loop travelling along the 1-handle appears as a meridian of the corresponding dotted unknot. Both 1-handle notations are shown in Figure 13.2.

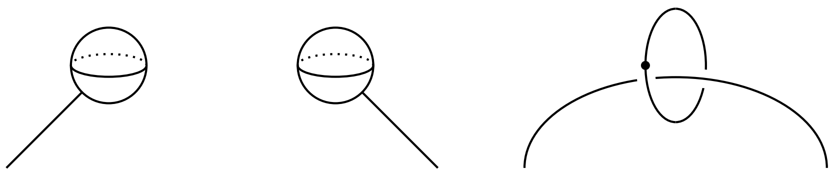


Figure 13.2. Left: The ball notation for a 1-handle attachment. Right: The dotted circle notation for a 1-handle attachment. In both diagrams, a curve that runs over the 1-handle is shown in black.

We will principally use the dotted circle notation for 1-handles. The two distinct notations for 1-handles are related as follows. Given two solid balls in $\mathbb{R}^3$ denoting the attaching region of a 1-handle, isotope them to come close to each other, flattening them into parallel thickened discs in the process. The pair of discs is replaced by a dotted unknot parallel to their boundary.

Consider a Kirby diagram consisting of a single dotted unknot U in S^3. We know from before that the corresponding 4-manifold X is diffeomorphic to $S^1 \times D^3$, which has boundary $S^1 \times S^2$. We take a moment now to see this boundary in the Kirby diagram itself. By our construction, there exists a pushed in disc $\Delta \subseteq D^4$ bounded by U and $X = \overline{D^4 \setminus \nu\Delta}$, where $\nu\Delta$ is a closed tubular neighbourhood of Δ. So the boundary of X decomposes as

$$\partial X = \overline{S^3 \setminus \nu U} \cup \Delta \times S^1,$$

glues along the boundary $\partial \nu U$. Since $\Delta \times \{1\}$ is bounded by a 0-framed longitude of the unknot, the boundary is identified with the 0-framed Dehn surgery on S^3 along the unknot U, which is diffeomorphic to $S^1 \times S^2$. Note in particular that the boundary is the union of the complement $S^3 \setminus U$ and the surgery solid torus $\Delta \times S^1$. A tubular neighbourhood of a meridian of the dotted circle is isotopic in the boundary to the surgery solid torus $\Delta \times S^1$.

The boundary of a 4-manifold given by a Kirby diagram consisting of a dotted unlink has a similar decomposition, namely into the complement of the unlink in S^3 and a collection of surgery solid tori. In particular, a Kirby diagram consisting solely of a dotted unlink can be converted to a surgery diagram for its boundary by erasing all the dots and decorating all the components with zeros. Moreover, tubular neighbourhoods of meridians of the dotted unlink are isotopic in the boundary to the surgery solid tori.

13.2.2. Attaching 2-handles. By convention, the 2-handles in a handle decomposition are attached after all the 1-handles have been attached to the 0-handle. So we consider attaching 2-handles to $\partial(\natural^n S^1 \times D^3) = \#^n S^1 \times S^2$, where possibly $n = 0$ when there are no 1-handles. Recall that a 4-dimensional 2-handle is a copy of $D^2 \times D^2$. The attaching region is a copy of $S^1 \times D^2$, and the attaching sphere is a copy of S^1. The isotopy class of the attaching map is determined by a knot K in $\#^n S^1 \times S^2$ corresponding to the image of the attaching sphere and a longitude of K corresponding to the image of $S^1 \times \{\text{pt}\}$ for some pt $\in \partial D^2$. Note that this longitude bounds a parallel push-off of the core of h in $\partial_+ h$, where h is the 2-handle. In particular, the pushed off disc is disjoint from the core of h. The choice of longitude of K determines the framing of the normal bundle of K. As mentioned earlier, the choices of framing are in bijection with the elements of $\pi_1(O(2), \mathrm{Id}) \cong \pi_1(SO(2)) \cong \mathbb{Z}$. When we attach a 2-handle to S^3, namely when there are no 1-handles in the decomposition, the Seifert framing, corresponding to the Seifert longitude, provides a specified framing of the normal bundle of K. All other framings of the normal bundle of K can then be compared to the Seifert framing, and thus correspond to well defined integers. As a result, a 2-handle attachment to S^3 is indicated by a knot decorated by an integer.

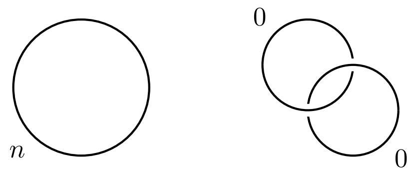


Figure 13.3. Left: A Kirby diagram for the D^2-bundle over S^2 with Euler number n. Right: A Kirby diagram for $S^2 \times S^2$ with an open 4-ball removed.

In the presence of 1-handles, framings of 2-handles are more difficult to describe. However, in the dotted circle notation for 1-handles, the 2-handle attaching spheres form a collection of knots in the complement of a dotted unlink in S^3. In particular, the 2-handle attaching spheres form a link in S^3, and we may use the Seifert framing as before.

A Kirby diagram for a 4-manifold built by attaching 1-handles and 2-handles to a 0-handle consists of a link in S^3, usually depicted in $\mathbb{R}^3$, such that each component is decorated either by an integer or a dot, and the dotted components form an unlink. In practice, all the 2-handles we encounter will be 0-framed, that is they will have the Seifert framing.

Consider the effect of adding a 2-handle on the boundary of a 4-manifold. Since the boundary of a 2-handle decomposes as $\partial(D^2 \times D^2) = S^1 \times D^2 \cup D^2 \times S^1$, the effect is to remove a solid torus neighbourhood of the image of the attaching circle from the boundary, and then add in another solid torus, via some diffeomorphism of the boundary torus. In other words, the effect on the boundary is a Dehn surgery. By the previous section, a Kirby diagram for a 4-manifold consisting of a link in S^3 decorated with dots and zeros can be converted to a surgery diagram for the boundary by replacing all the dots by zeros.

13.2.3. Attaching and tip regions. Our goal in this chapter is to provide Kirby diagrams for gropes, capped gropes, towers, and capped towers. The information we give will suffice to build Kirby diagrams of generalised towers. Recall that these are manifolds with corners, equipped with specified attaching and tip regions. The attaching and tip regions of a generalised tower are submanifolds of the boundary consisting of framed tubular neighbourhoods of simple closed curves, and the corners of a generalised tower are formed precisely by the boundaries of these tubular neighbourhoods. We will record the attaching and tip regions of a generalised tower in its Kirby diagram by drawing the cores of the attaching and tip regions, that is the attaching and tip circles, in the boundary, in fact in the complement in S^3 of the link comprising the Kirby diagram. We must also specify the framing of these curves. However, in practice the framing will always be the Seifert framing, as we explain in Section 13.4. Finally, we must introduce the appropriate corners (see Remark 12.2).

13.3. Kirby calculus

Handle decompositions, and thus Kirby diagrams, for a given 4-manifold are not unique. However, it follows from a result of Cerf [**Cer70**] that any two handle decompositions for a given fixed compact manifold are related by a sequence of *handle slides*, *birth/cancellation* of handle pairs, and isotopies within levels. These appear as purely diagrammatic moves within Kirby diagrams, and the art of using these moves to pass between Kirby diagrams is known as *Kirby calculus*. Isotopies within levels appear within a Kirby diagram as isotopies of the attaching regions of handles. We describe handle slides and the birth/cancellation of handle pairs next.

13.3.1. Handle slides. A handle slide is a special case of an isotopy of the attaching region of a handle. We will only consider the case of sliding 2-handles, since that is all we shall need, but one may slide handles of other indices as well. Given two 2-handles h_1 and h_2 attached to some 4-manifold X along disjoint simple closed curves in ∂X, the *handle slide* of h_1 over h_2 consists of isotoping the attaching sphere of h_1 in $\partial(X \cup h_2)$, by pushing across and over $\partial_+ h_2$, until it returns to ∂X. In other words, the attaching sphere of h_1 moves through the belt sphere of h_2.

Within a Kirby diagram, suppose that the attaching spheres for h_1 and h_2 are given by the knots K_1 and K_2 with framings f_1 and f_2 respectively. Let K_2' be the f_2-framed longitude of K_2, with linking number $\ell\mathrm{k}(K_2, K_2') = f_2$. Recall that a *band sum* of knots is similar to a connected sum, where the band used for summing need not be unknotted nor unlinked from the knots in question. In the Kirby diagram, the handle slide of h_1 over h_2 replaces the knot K_1 by the knot $\overline{K}_1$, which is some choice of band sum of K_1 with K_2', equipped with the framing $f_1 + f_2 \pm 2\,\ell\mathrm{k}(K_1, K_2')$.

The choice of sign in the framing formula depends on the orientations of the handles used. However, we omit these details since we will only slide 2-handles with trivial linking number in the sequel. See [**GS99**] for the general case. For us, it suffices to know that when sliding 0-framed 2-handles with trivial linking numbers over one another, the knots change by band sums and the framings remains trivial; see Figure 13.4.

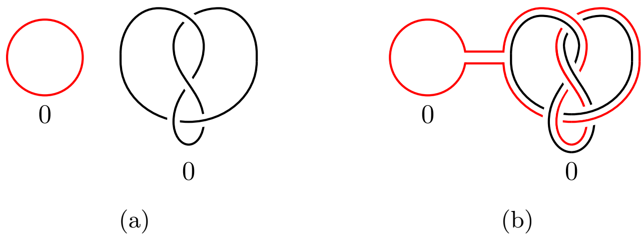


Figure 13.4. (a) Two 0-framed 2-handles attached to a 0-handle. (b) The red handle has been slid over the black handle.

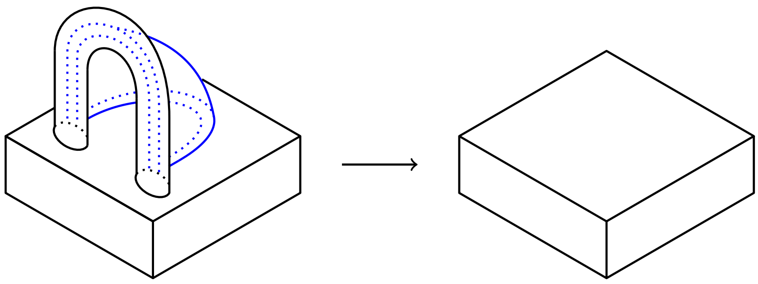


Figure 13.5. A cancelling 1- and 2-handle pair in three dimensions.

13.3.2. Handle cancellation. In three dimensions, the canonical picture for a handle cancellation is the igloo melting back into the arctic (see Figure 13.5), where a 3-dimensional 1-handle (the black archway) and a 3-dimensional 2-handle (the blue roof and walls) are cancelled, by isotoping along the roof and down into to the base. In general, in any dimension, a k-handle h_k and a $k+1$-handle h_{k+1} may be cancelled when the attaching sphere of h_{k+1} intersects the belt sphere of h_k transversely and at precisely one point. A handle cancellation means that we delete both h_k and h_{k+1} from the handle decomposition after moving any other handles interacting with h_k and h_{k+1} out of the way. Such a cancelling pair of handles may also be added to a handle decomposition at will.

We restrict ourselves to the cancellation of 1-handles and 2-handles, referring again to [**GS99**] for the general case. By our previous statement, when a Kirby diagram contains a dotted circle isotopic to the meridian of the attaching circle of a 2-handle, then we may delete both the 2-handle attaching circle and the dotted circle from the diagram. In general, whenever we see the attaching circle for a 2-handle h_2 linking a dotted circle corresponding to a 1-handle h_1 geometrically once, we may perform handle slides over h_2 to move aside all other 2-handle attaching circles passing through h_1 to reduce to the previous case (see Figure 13.6).

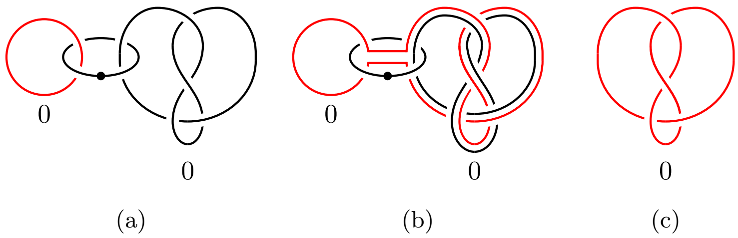


Figure 13.6. Cancellation of 1- and 2-handles in a Kirby diagram. (a) Two 2-handles and a 1-handle attached to a 0-handle. (b) The red 2-handle is slid over the black 2-handle. (c) The black 2-handle and the 1-handle have been cancelled.

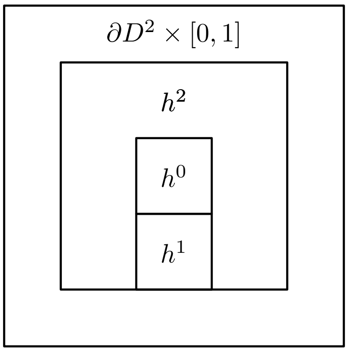


Figure 13.7. Decomposing a disc D^2 into handles relative to its boundary where h^i denotes an i-handle.

13.3.3. Plumbing. We finish this section by describing Kirby diagrams for plumbed and self-plumbed 2-handles. This will be key in our construction of Kirby diagrams for capped gropes, towers, and capped towers.

We recall the definition of plumbing from Section 12.1 for the convenience of the reader. To plumb together two 2-handles, namely two copies of $D^2 \times D^2$, we take small closed disc neighbourhoods D and E in the interior of the core of each 2-handle and then identify $D \times D^2$ and $E \times D^2$ by setting $(x, y) \in D \times D^2$ to be equal to $(y, x) \in E \times D^2$ for a positive plumbing, or equal to $(-y, -x)$ for a negative plumbing. (Recall that we smooth corners in the plumbing regions, so that corners only occur where they did in the original 2-handles prior to plumbing.) For a self-plumbing of a single 2-handle $D^2 \times D^2$, the discs D and E are required to be disjoint closed disc neighbourhoods in the interior of the core $D^2 \times \{0\}$. It is straightforward to see that the result of a single self-plumbing on a 2-handle $D^2 \times D^2$ is diffeomorphic to $S^1 \times D^3$. Indeed the result of self-plumbing n times on a 2-handle $D^2 \times D^2$ is diffeomorphic to $\natural^n S^1 \times D^3$.

In the description above, note that the subdisc D within the core of the first 2-handle is identified with the cocore of the second 2-handle. This is the key observation motivating the Kirby diagram for plumbed 2-handles. Let h and h' be two 2-handles attached to a 4-manifold X. The core disc of h admits a handle decomposition relative to its boundary consisting of a 0-handle, a 1-handle, and a 2-handle as shown in Figure 13.7. Note that these are 2-dimensional handles. The 0-handle in this figure is a subdisc D in the core of h. The product of this handle decomposition with D^2 yields a handle decomposition of h relative to its attaching region. Suppose we wish to plumb together h and h'. Then, first detach h from X.

Identify the 0-handle in the decomposition of the core of h mentioned earlier with a cocore of h'. Each D^2 fibre of a point in the 0-handle is identified with a parallel copy of a small closed subdisc in the core of h' (corresponding to the subdisc E from earlier). Then reattach the rest of h by attaching the remaining 1-handle and 2-handle. The result is the space obtained by plumbing together h and h'.

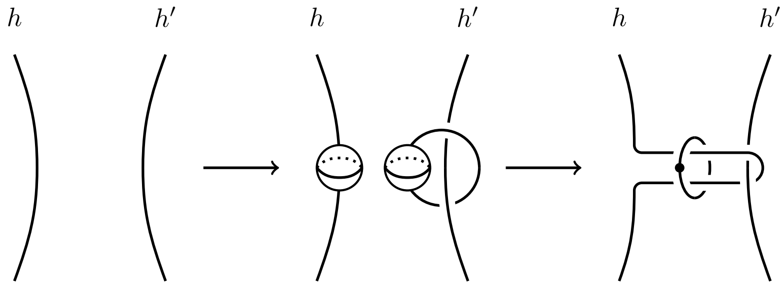


Figure 13.8. Plumbing together the 2-handles h and h'. The framings of h and h' remain the same. The result is shown using both notation conventions for 1-handles.

Since the cocore of a 2-handle is isotopic to a meridian of its attaching circle, the effect of plumbing two 2-handles in a Kirby diagram is to introduce a clasp between the corresponding attaching circles via an unknotted, untwisted band running directly through a 1-handle, as shown in Figure 13.8. The framings of the handles remain the same. Note that if the handles h and h' are oriented, the sign of the clasp will correspond to the sign of the intersection between the cores of h and h' created by the plumbing operation.

A more careful analysis is necessary when performing a self-plumbing. As noted earlier, now the discs D and E are subdiscs of the core disc $D^2 \times \{0\}$ in $D^2 \times D^2$. Once the identification of $D \times D^2$ with $E \times D^2$ has taken place, switching the factors as usual, the boundary circles of D and E appear as an (oriented) Hopf link in the S^3 formed as the boundary of $D \times D^2 \equiv E \times D^2$. The sign of the Hopf link corresponds to the sign of the plumbing performed. The boundary of $D^2 \times \{0\}$ retracts to the band sum of the two components of the Hopf link, via an untwisted band travelling around a 1-handle (compare with Figure 1.6). As before the framing of the handle remains the same.

Pictorially, the process for self-plumbing a single handle in a Kirby diagram agrees with the process for plumbing together two distinct handles (see Figure 13.9). In the self-plumbing case, the dotted circle corresponding to the 1-handle in Figure 13.8 can be isotoped to form a Whitehead double of the meridian of the attaching circle of h, as shown in Figure 13.9. The sign of the self-plumbing and the sign of the clasp match. Note that unlike the case where we plumbed two distinct 2-handles together, the sign of the self-plumbing is fixed regardless of the choice of orientation of the 2-handle being self-plumbed.

We take a moment now to detect the core discs of the plumbed 2-handles in these Kirby diagrams. Observe in the rightmost panel of Figure 13.8 that we may perform a crossing change to unlink h and h'. The trace of this homotopy, in a collar $S^3 \times [0, \varepsilon]$ of $S^3 = \partial D^4$, glued to the cores of the handles h and h' forms the cores of the original handles before plumbing. Similarly, in Figure 13.9, there is a homotopy, consisting of a crossing change, which undoes the clasp in the attaching circle of the handle h, regardless of the sign of the clasp. The trace of this homotopy along with the core of the handle forms the core of the original handle before plumbing.

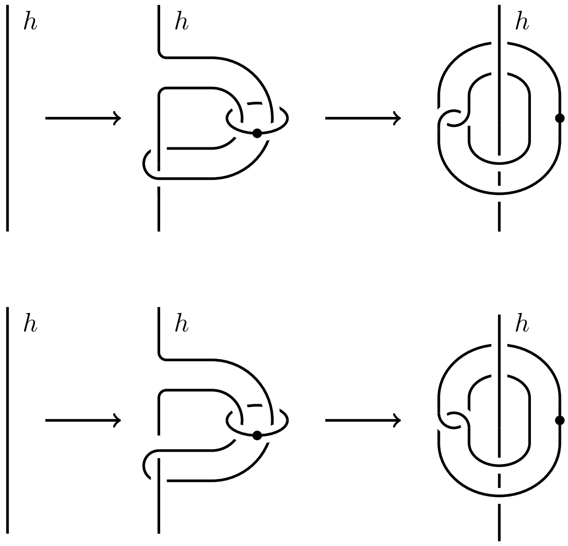


Figure 13.9. Diagrams for self-plumbing a 2-handle h. Both a positive (top) and a negative (bottom) self-plumbing are shown. The final diagram in each panel shows the result of an isotopy.

13.4. Kirby diagrams for generalised towers

Now we begin to construct Kirby diagrams for gropes, capped gropes, towers, and capped towers. We show first how to construct Kirby diagrams for surface, disc, and cap blocks, and then how to combine these diagrams as necessary.

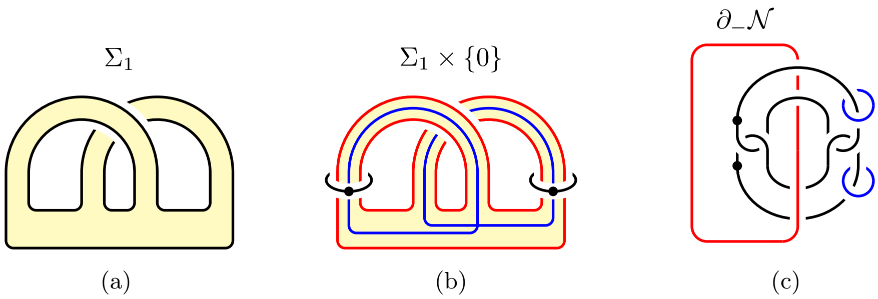


Figure 13.10. (a) The surface Σ_1 (yellow) is produced by attaching a pair of 2-dimensional 1-handles to a 2-dimensional 0-handle. (b) By taking the product of the handle decomposition in (a) with D^2, we obtain a (4-dimensional) handle decomposition of $\Sigma_1 \times D^2$. The handle decomposition consists of the pair of 1-handles denoted by the two dotted circles, attached to a 0-handle. The surface $\Sigma_1 \times \{0\}$ is shown in yellow with red boundary. A symplectic basis of curves for $H_1(\Sigma_1 \times \{0\})$ is shown in blue. (c) The result of an isotopy of the Kirby diagram in (b). The red circle is the image of the boundary of $\Sigma_1 \times \{0\}$. This is a Kirby diagram for a surface block $\mathcal{N}$ with genus one. The attaching region $\partial_- \mathcal{N}$ is a trivially framed tubular neighbourhood of the red circle, while the tip region consists of trivially framed tubular neighbourhoods of the blue circles.

13.4.1. Surface blocks. Let Σ_g denote the abstract oriented surface with genus g and a single boundary component. Consider Σ_1 to start. Note that Σ_1 is constructed by attaching a pair of 2-dimensional 1-handles to a 2-dimensional 0-handle, as in Figure 13.10(a). Taking the product with D^2 produces a handle decomposition of $\Sigma_1 \times D^2$ consisting of a pair of 4-dimensional 1-handles attached to a 4-dimensional 0-handle, in which $\Sigma_1 \times \{0\}$ is still visible, as in Figure 13.10(b). An isotopy shows that the boundary of $\Sigma_1 \times \{0\}$ and the dotted circles denoting the 1-handles form the Borromean rings in S^3. Moreover, a symplectic basis of curves for $H_1(\Sigma_1; \mathbb{Z})$ is isotopic to meridians of the dotted circles. Figure 13.10(c) thus gives a Kirby diagram for a surface block with genus one, where the core of the attaching region is denoted in red and the cores of the tip regions are denoted in blue. The spine of the block is pictured in Figure 13.10(b). As a sanity check, note that as expected $\Sigma_1 \times D^2$ is diffeomorphic to $S^1 \times D^3 \natural S^1 \times D^3$.

To understand the framing of the attaching and tip regions, we recall the definition of the abstract surface block from Chapter 12 in greater detail. Begin with a standard surface in S^3 with a single boundary component, thicken first within S^3, and then again by multiplying the model with $[0, 1]$. The framing of the attaching circle is obtained from the tangent vectors to the two thickenings. Since the first thickening takes place within S^3, the framing is the Seifert framing. The framing of the tip regions is given by the surface framing and the second thickening direction. From Figure 13.10(b), this is also seen to be the Seifert framing.

More generally, one can follow the same procedure starting with a decomposition of Σ_g into $2g$ 1-handles attached to a 0-handle, for any $g \geq 1$. In this case, we obtain a Kirby diagram for a surface block of genus g as indicated in Figure 13.11. Again the attaching region is given by a trivially framed neighbourhood of the red circle, while the tip region consists of trivially framed neighbourhoods of the blue circle.

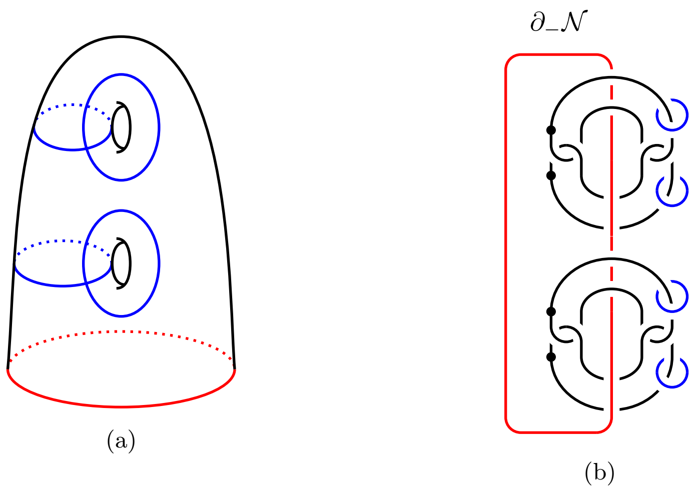


Figure 13.11. (a) The abstract surface Σ_2 with boundary shown in red and a symplectic basis of curves for $H_1(\Sigma_2; \mathbb{Z})$ shown in blue. (b) A Kirby diagram for the surface block $\mathcal{N}$ with genus two. The attaching region $\partial_- \mathcal{N}$ is a trivially framed neighbourhood of the red circle. The tip region consists of trivially framed neighbourhoods of the blue circles.

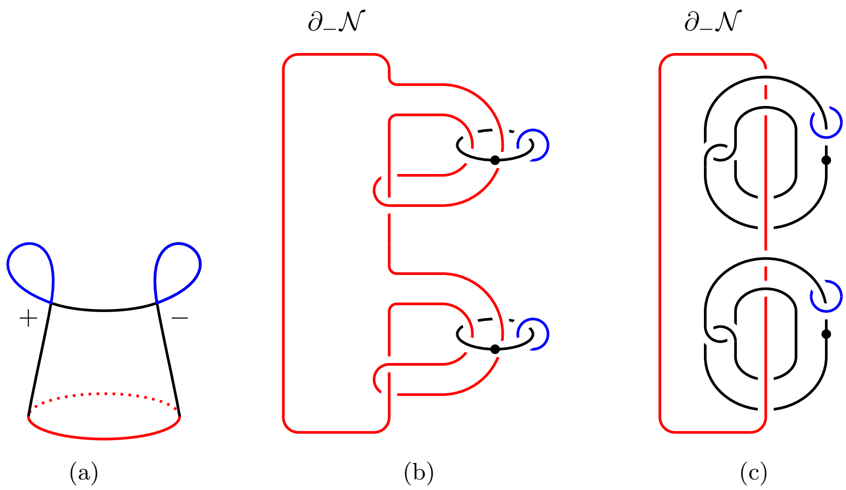


Figure 13.12. (a) A schematic picture of the 2-dimensional spine of a disc block with one positive and one negative self-plumbing. (b) A Kirby diagram for the disc block indicated in (a). The attaching region $\partial_-\mathcal{N}$ is a trivially framed neighbourhood of the red curve and the tip region consists of trivially framed neighbourhoods of the blue curves. Note that if we forget the tip regions, this is also a Kirby diagram for a cap block. (c) The result of an isotopy on the diagram in (b).

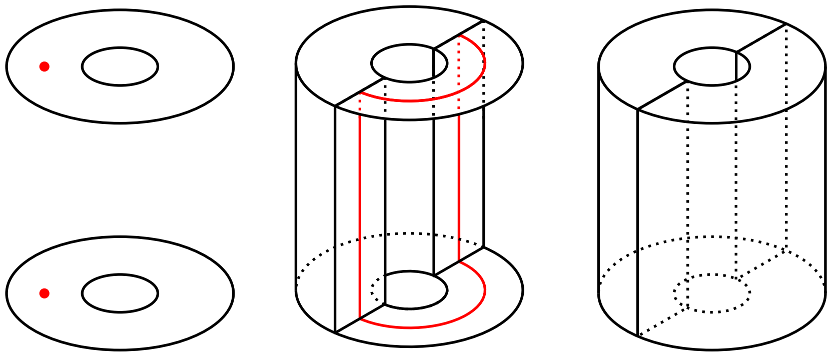
13.4.2. Disc and cap blocks. Recall that disc and cap blocks both consist of self-plumbed thickened discs. In the language of this chapter, these are self-plumbed 2-handles. They are not yet attached to any 4-manifold, but the attaching region records how they should be attached. From Section 13.3.3, we see that Figure 13.12 depicts a Kirby diagram for a self-plumbed 2-handle. In particular, the 4-manifold is the boundary connected sum $\natural^n S^1 \times D^3$ for some n, as expected. In a disc block the number of positively and negatively clasped dotted circles must match up, and the meridians of these dotted circles form the tip regions. In a cap block, there is no restriction on the signs of the clasps of the dotted circles and there are no tip regions.

Next we describe the spine of the disc and cap blocks. Let $\mathcal{N}$ be a disc or cap block with underlying 4-manifold N. Following Section 13.3.3, the spine of $\mathcal{N}$ is an immersed disc in N with transverse self-intersections in the interior, seen in Figure 13.12(b) as follows. There is a homotopy of the attaching circle in the complement of the dotted circles, consisting of crossing changes, taking it to an unknotted circle split and unlinked from the dotted circles. That is, the result of the homotopy bounds a disc away from the dotted circles. The trace of the homotopy glued to this disc is the spine of $\mathcal{N}$.

It remains only to detect the framing of the attaching and tip regions. From the discussion in Section 13.3.3 it follows immediately that the tip regions of a disc block have the trivial Seifert framing; this can also be seen by locating the Whitney circles and accessory circles on the spine and then pushing off (see Figure 12.2). The tip region of a cap block is empty by definition. Recall from Chapter 12 that the framing on the attaching region of a disc or cap block is that of the attaching region $S^1 \times D^2$ of $D^2 \times D^2$ before plumbings are introduced. We assert that this implies that the attaching circle also has the trivial Seifert framing. This follows directly from the fact that if the tip regions are capped off with 2-handles, using the tip region framing, the result is the (unplumbed) 2-handle with the canonical, trivial framing of its attaching region. Note, in particular, that the Seifert framing on the attaching region coincides with the framing induced by the normal bundle of the spine of $\mathcal{N}$ if and only if the spine has equal numbers of positive and negative double points. Indeed, from the description of the spine in the previous paragraph we see that the normal bundle of the spine of the disc block in Figure 13.12(b) induces the blackboard framing.

13.4.3. Stages. Recall that a stage is a collection of blocks, where each surface block and each disc block is disjoint from all other blocks, but cap blocks may intersect one another in arbitrary plumbings. Since a 4-manifold given by a Kirby diagram is necessarily connected, surface and disc stages are represented by a collection of disjoint Kirby diagrams. Given a connected stage consisting of several cap blocks intersecting one another in plumbings, we may form a Kirby diagram by following the recipe of Section 13.3.3. However, this will not be relevant for us and we leave it to the reader.

13.4.4. Generalised towers. As remarked above, Kirby diagrams necessarily represent connected 4-manifolds. Thus a disconnected 4-manifold is represented by a disjoint union of Kirby diagrams and we only concern ourselves henceforth with Kirby diagrams for connected generalised towers.



(a) (b) (c)

Figure 13.13. (a) Two copies of $S^1 \times D^2$ to be identified. The third dimension is suppressed. The attaching sphere for a 1-handle is indicated in red. (b) After attaching a 1-handle. The attaching circle for a 2-handle is shown in red. (c) A copy of $S^1 \times D^2 \times [0,1]$ has been constructed.

Recall that a generalised tower is built by stacking stages on top of each other by identifying tip and attaching regions. Each connected component of the tip and attaching regions is a copy of a solid torus $S^1 \times D^2$ in the boundary of a 4-manifold. Two such solid tori in two distinct 4-manifolds may be identified by attaching a copy of $S^1 \times D^2 \times [0, 1]$, where the boundary components $S^1 \times D^2 \times \{0, 1\}$ are mapped to the solid tori to be identified. Note that $S^1 \times D^2 \times [0, 1]$ decomposes relative to $S^1 \times D^2 \times \{0, 1\}$ as a union of a 1-handle and a 2-handle, as indicated by Figure 13.13. Thus the identification of two solid tori may be accomplished by adding a 1-handle and a 2-handle. When the solid tori being identified are in distinct 4-manifolds, the 1-handle attachment merely connects the 0-handles of the two 4-manifolds and need not be drawn. More precisely, we identify the union of the two 0-handles and the 1-handle with the new 0-handle by drawing the two

Kirby diagrams side by side. In our case, the 2-handle attaching circle will wrap around the longitude of a component of a tip region, then over the 1-handle, then around the longitude of a component of the attaching region, then back over the 1-handle.

In practice, if a block $\mathcal{N}_2$ is to be stacked on top of a block $\mathcal{N}_1$, we place the Kirby diagrams next to each other, take the connected sum of the curves denoting the corresponding tip and attaching regions, and declare the resulting curve to be a 2-handle attaching circle. Since the attaching and tip regions were trivially framed, the new curve is also trivially framed. The result is a Kirby diagram for the manifold produced by stacking $\mathcal{N}_2$ on top of $\mathcal{N}_1$ as desired. We illustrate this procedure with a number of pictures.

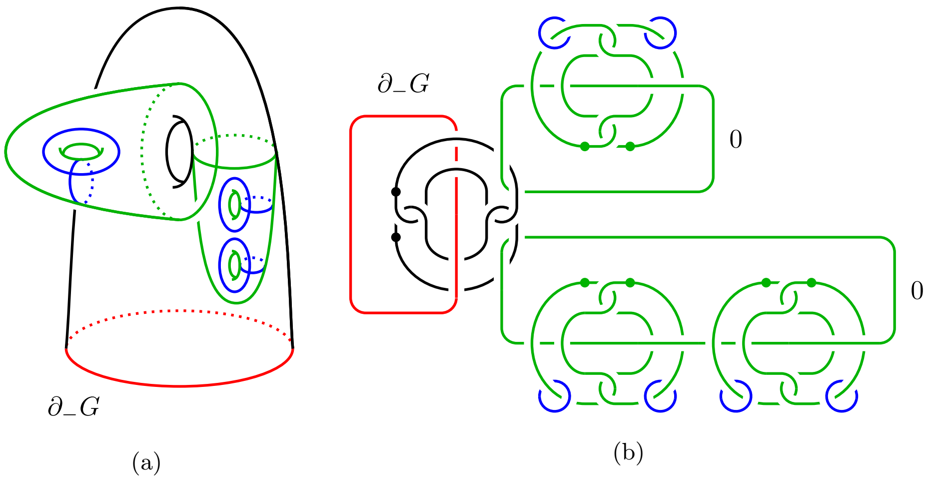


Figure 13.14. (a) The 2-dimensional spine of a height two grope G. Tip circles are shown in blue. (b) A Kirby diagram for the grope indicated in (a). The attaching region $\partial_- G$ is a trivially framed neighbourhood of the red circle. The second stage surfaces are shown in green in both panels. The tip region consists of trivially framed neighbourhoods of the blue circles.

Using the procedure just described, the reader should now be able to draw Kirby diagrams for arbitrarily complicated generalised towers. However, as exemplified by Figure 13.16, these diagrams rapidly increase in complexity, as the number of stages increases. We will soon see how to simplify these diagrams.

13.5. Bing and Whitehead doubling

We recall some convenient terminology that will be used throughout the rest of the book. Let L be a link in S^3. Let νL denote a collection of pairwise disjoint closed tubular neighbourhoods of the components of L. Replace each component of νL by the solid torus in Figure 13.17(a), mapping the standard longitude of the solid torus in the figure to the 0-framed longitude of the corresponding component of L. The resulting 3-manifold is diffeomorphic to S^3 and the link formed by the collection of images of the red circles is called the *Bing double* of L.

The *Whitehead double* of L is produced in exactly the same way, except using the pattern in Figure 13.17(b). Note that there is a choice of sign for the clasp. The clasp shown corresponds to *negative Whitehead doubling*, while the other clasp corresponds to *positive Whitehead doubling*. When we omit to mention the sign of

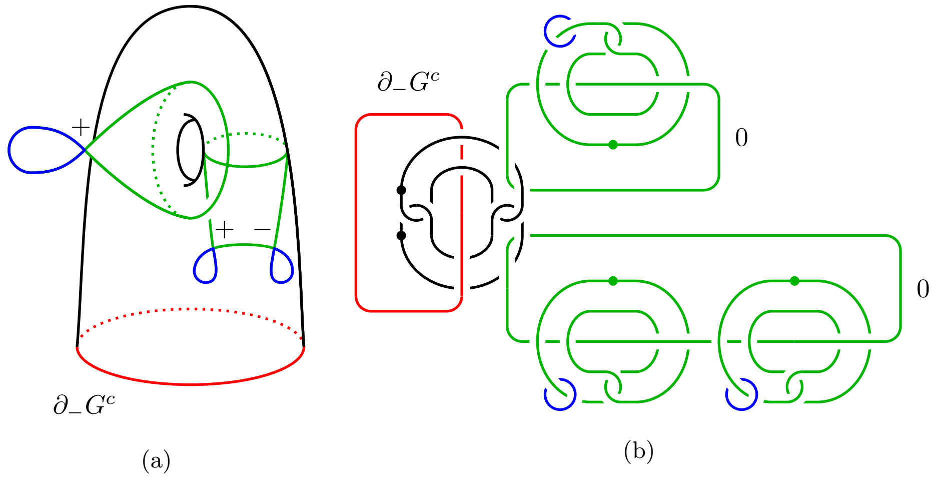


Figure 13.15. (a) The 2-dimensional spine of a height one capped grope G^c. (b) A Kirby diagram for the capped grope indicated in (a). The attaching region $\partial_- G^c$ is a trivially framed neighbourhood of the red circle. The caps are shown in green in both panels. The tip region consists of trivially framed neighbourhoods of the blue circles.

Whitehead doubling, it should be assumed that the sign is irrelevant, and either sign may be chosen.

A *ramified Bing double* of L is any link obtained by first replacing L by a link L' consisting of a nonzero number of parallel copies of each component of L, and then taking the Bing double of L'. The number of parallel copies taken of distinct components of L need not be the same. Of course, the Bing double is the result of ramified Bing doubling where only one parallel copy of each component of L is taken.

An *iterated, ramified Bing double* of L is the result of performing several iterations of ramified Bing doubling operations on L. Note that the ramification parameters need not be constant across components nor across iterations. That is, any positive number of parallel copies may be taken for any component and at any stage of the iteration.

Similarly, we define a *ramified Whitehead double* of L and an *iterated ramified Whitehead double* of L. The signs of the clasps for the components may vary freely.

Finally, an *iterated, mixed, ramified Bing-Whitehead double* of L is the result of a concatenation of ramified Bing doubling and ramified Whitehead doubling operations on L. We again impose no conditions at this stage on the signs of the Whitehead doubles, and ramification parameters can vary freely across components and stages of the iteration.

Observe that this can also be used to define iterated, ramified Bing and Whitehead doubling on a link in a solid torus or on a sublink of a link in S^3. In the former case, we embed the solid torus in S^3 in the standard unknotted manner. For example, the link in Figure 13.23, ignoring all dots and decorations, can be considered to be the result of iterated mixed ramified Bing-Whitehead doubling of the core of the solid torus given by the complement of the red circle, or alternatively the result of performing iterated mixed ramified Bing-Whitehead doubling on one of the two components of the Hopf link in S^3.

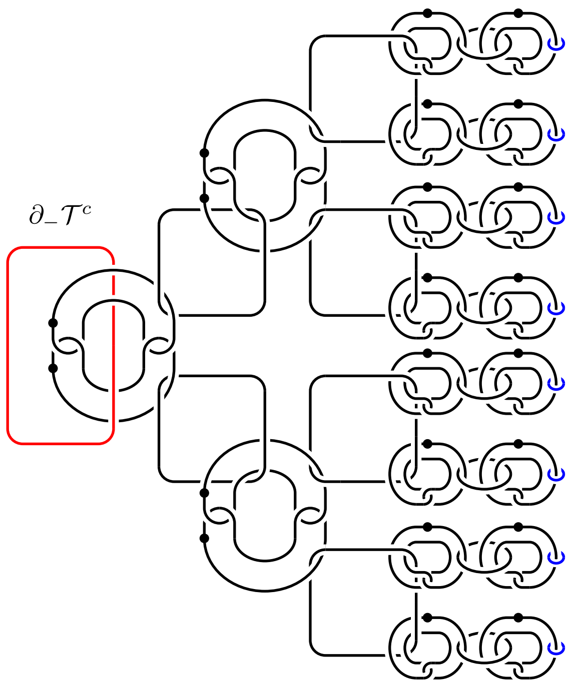
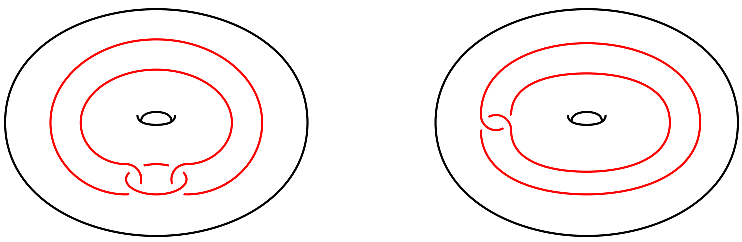


Figure 13.16. A Kirby diagram of a 1-storey capped tower $\mathcal{T}^c$. All solid black curves without dots are 0-framed by default. The attaching and tip regions are trivially framed neighbourhoods of the red and blue circles respectively, as usual. We invite the reader to draw an abstract picture of the 2-dimensional spine of $\mathcal{T}^c$.



(a) (b)

Figure 13.17. (a) The pattern for Bing doubling. (b) The pattern for (negative) Whitehead doubling.

Definition 13.1 (Mixed ramified Bing-Whitehead link). A link in the solid torus $S^1 \times D^2$ obtained by performing iterated mixed ramified Bing-Whitehead doubling on the core $S^1 \times \{0\}$ is called a *mixed ramified Bing-Whitehead link*.

13.6. Simplification

So far we have produced Kirby diagrams involving both 1-handles and 2-handles. Some of these were quite complicated. However, note that the attaching circle for every 2-handle in these Kirby diagrams of a generalised tower occurs as the meridian of a dotted circle. As a result, we may cancel the corresponding 1- and 2-handle pair from the diagram. However, we must perform some preliminary handle slides and keep track of how the rest of the diagram changes. This is the goal of this section. We will see that every grope, capped grope, tower, and capped tower has a handle decomposition consisting only of 1-handles, or equivalently a Kirby diagram consisting only of dotted circles. The 4-manifolds are therefore all diffeomorphic, after smoothing corners, to a boundary connected sum of copies of $S^1 \times D^3$. Of course, we will still have to keep track of the attaching and tip regions.

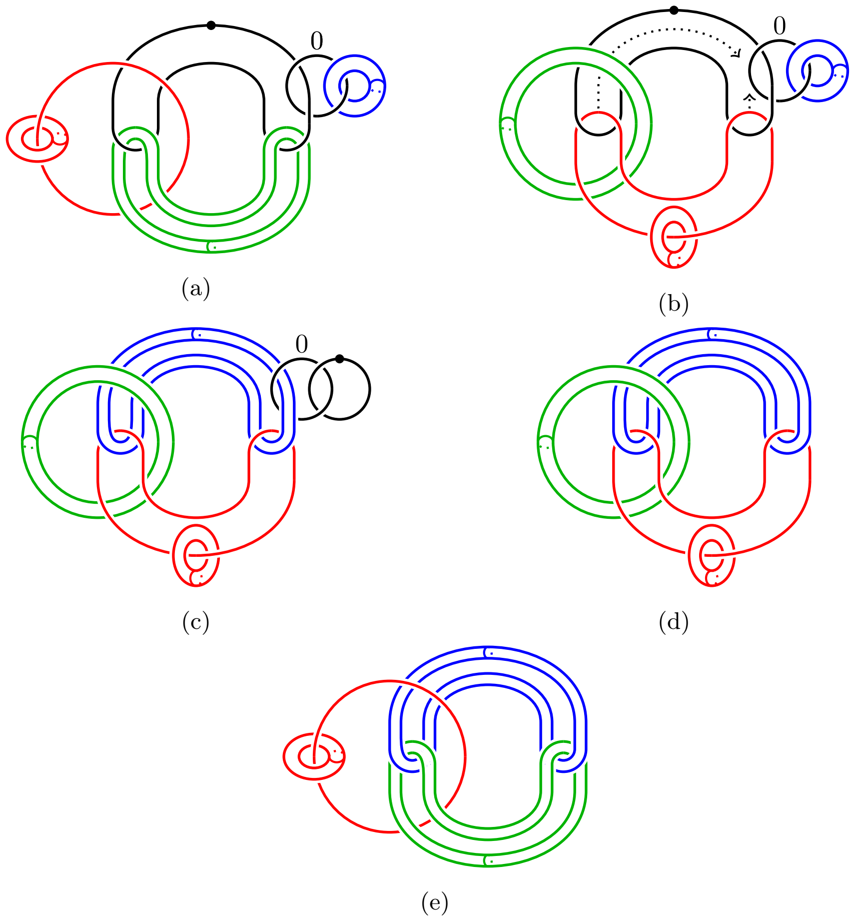


Figure 13.18. Simplification for Bing doubles. The red, blue, and green solid tori represent portions of some larger Kirby diagram. The first two diagrams are isotopic. Perform two handle slides as shown followed by another isotopy to obtain the third diagram. Cancel the 1-/2-handle pair to obtain the fourth diagram. Another isotopy yields the fifth and final diagram.

13.6.1. Bing doubles. Consider first the situation shown in Figure 13.18(a). We see a dotted circle and the attaching circle of a 2-handle (both shown in black) linking precisely once geometrically. This is the pair we wish to cancel. The red circle may either be the 0-framed attaching circle for a 2-handle or represent the attaching region. All other ingredients of the Kirby diagram, including the tip and attaching regions, are contained within the blue, green, and red solid tori; compare with Figures 13.14, 13.15, and 13.16. The solid tori will move around in the subsequent steps, with the understanding that there is no twisting or modification within them.

Observe first that the black dotted circle, the red circle, and the green solid torus, form a copy of the Borromean rings. Since the Borromean rings are symmetric in their components, there is an isotopy of the diagram exchanging the rôles of the red curve and the dotted circle. The result of this isotopy is shown in Figure 13.18(b).

Slide the red curve over the black 2-handle attaching circle twice as indicated in Figure 13.18(b). This is either an isotopy, when the red circle indicates the attaching region, or a handle slide, when the red circle is itself a 2-handle attaching circle. In both cases, the framing of the neighbourhood of the red circle remains trivial, as desired, since the black attaching circle is 0-framed and has trivial linking number with the red circle (computing framing coefficients after handle slides was explained in Section 13.3.1). The result is shown in Figure 13.18(c).

Now the only curve passing through the dotted circle is the black 2-handle attaching circle, allowing us to cancel both of these from the diagram. This results in Figure 13.18(d). Finally observe that the red curve and the blue and green tori in Figure 13.18(d) form a copy of the Borromean rings, so there is an isotopy to Figure 13.18(e).

In summary, a Kirby diagram such as Figure 13.18(a) may be replaced by the diagram in Figure 13.18(e). As desired, the 1- and 2-handle pair shown in black has been cancelled.

13.6.2. Whitehead doubles. We have a similar simplification for Whitehead doubles. The general picture here is shown in Figure 13.19(a). As before, we wish to cancel the dotted circle and the attaching circle of a 2-handle (both shown in black) which link precisely once geometrically. The red circle may either be the 0-framed attaching circle for a 2-handle or represent the attaching region. All other ingredients of the Kirby diagram, including the tip and attaching regions, are contained within the blue and red solid tori; compare with Figure 13.16.

Since the red curve and the dotted curve form a Whitehead link, there is an isotopy exchanging their rôles in the diagram, resulting in Figure 13.19(b). Slide the red curve over the black 2-handle attaching circle twice to obtain Figure 13.19(c), leaving the framing on the red curve unchanged. Then cancel the dotted circle and the 0-framed 2-handle to obtain Figure 13.19(d). Finally use the symmetry of the Whitehead link again to obtain Figure 13.19(e).

In summary, a Kirby diagram such as Figure 13.19(a) may be replaced by the diagram in Figure 13.19(e). As desired, the 1- and 2-handle pair shown in black has been cancelled.

13.6.3. Trees associated to generalised towers. Our goal is to provide simplified Kirby diagrams for generalised towers consisting only of 1-handles, along with the attaching and tip regions. This is obtained by taking a diagram as described in Section 13.4.4 (e.g. Figure 13.16), and iterating the simplification procedures of Sections 13.6.1 and 13.6.2. We describe the combinatorics of the resulting link by encoding the relevant data for a tower in a labelled tree, and then producing a link from the labelled tree.

A surface block $\mathcal{N}$ with spine Σ of genus g is associated to a tree $T_{\mathcal{N}}$ constructed as follows. Recall that the tip region of $\mathcal{N}$ arises from a symplectic basis $\mathcal{B}$ of simple closed curves for $H_1(\Sigma;\mathbb{Z})$. Begin with a root vertex associated to the attaching region. Attach g edges to the root. The result is called the α part of $T_{\mathcal{N}}$. Attach two edges to each new vertex. These new $2g$ edges form the β part of $T_{\mathcal{N}}$. The union of the α and β parts of $T_{\mathcal{N}}$ is the tree $T_{\mathcal{N}}$ (the α and β parts intersect in a collection of vertices). In addition, we prescribe a bijective correspondence between the set of leaves of $T_{\mathcal{N}}$ and the set of tip regions of $\mathcal{N}$, so that the edges in the β region arising from the same edge in the α region correspond to components of the

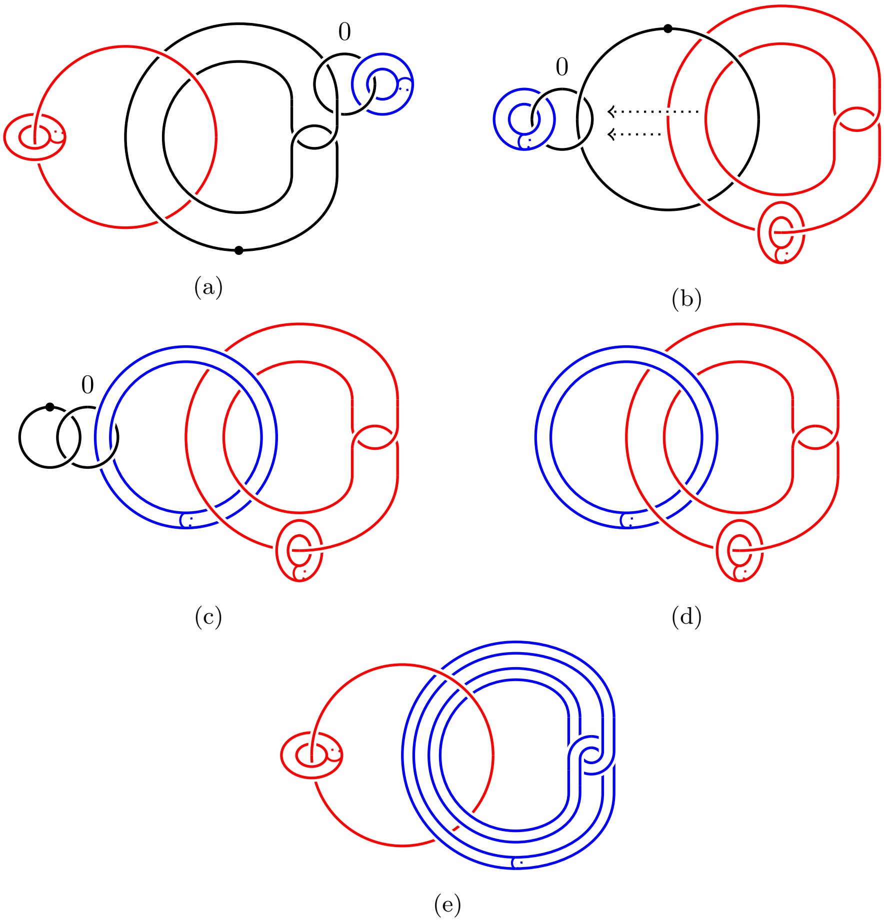


Figure 13.19. Simplification for Whitehead doubles. The red and blue solid tori represent portions of some larger Kirby diagram. The first two diagrams are isotopic. Perform two handle slides as shown followed by another isotopy to obtain the third diagram. Cancel the 1-/2-handle pair to obtain the fourth diagram. Another isotopy yields the fifth and final diagram.

tip region arising from a dual pair of curves in $\mathcal{B}$. (By convention the root of a tree is not a leaf, even when it has valence one.) Note that $T_{\mathcal{N}}$ can be embedded in Σ so that the leaves of $T_{\mathcal{N}}$ lie on the corresponding element of $\mathcal{B}$.

Given a disc or cap block $\mathcal{N}$, with p positive and n negative self-plumbings, there is a similar process to produce a tree $T_{\mathcal{N}}$. Begin again with a root vertex associated to the attaching region. Attach $p+n$ edges to the root. The result is called the α part of $T_{\mathcal{N}}$. Attach a new edge to each new vertex. These new $p+n$ edges form the β part $T_{\mathcal{N}}$. The union of the α and β parts of $T_{\mathcal{N}}$ is the tree $T_{\mathcal{N}}$. In addition, we prescribe a bijective correspondence between the set of leaves of $T_{\mathcal{N}}$ and the set of tip regions of $\mathcal{N}$, and we label the p vertices corresponding to the tip regions from a positive self-plumbing with a $+$, and the n vertices corresponding to the tip regions from a negative self-plumbing with a $-$. Note that $T_{\mathcal{N}}$ can be embedded in the spine of $\mathcal{N}$ so that the leaves of $T_{\mathcal{N}}$ lie on the corresponding double point.

It is now straightforward to construct trees for generalised towers. We restrict ourselves to disc-like generalised towers with no additional plumbings across distinct cap blocks since that is all we shall need later. By definition, such towers are constructed by identifying tip and attaching regions of the constituent blocks. By construction, given a surface, disc, or cap block, the root of the associated tree corresponds to the attaching region, and the leaves correspond bijectively to the tip regions. To form the tree associated to a generalised tower, identify the roots and leaves of the trees associated to the corresponding blocks. The component edges of the new tree inherit the designation into α and β regions. Moreover, record which edges came from surface blocks and which came from disc/cap blocks. When the generalised tower is homogeneous, we need only record the type of the corresponding stage. For a generalised tower $\mathcal{T} := G_0 \cup G_1 \cup \cdots \cup G_h$, denote the corresponding tree as $T_{\mathcal{T}} := T_{G_1} \cup \cdots \cup T_{G_h}$, where T_{G_i} is the forest corresponding to the stage G_i. Recall that G_0 denotes the attaching region.

The construction is best illustrated by examples, as in Figures 13.20 and 13.21.

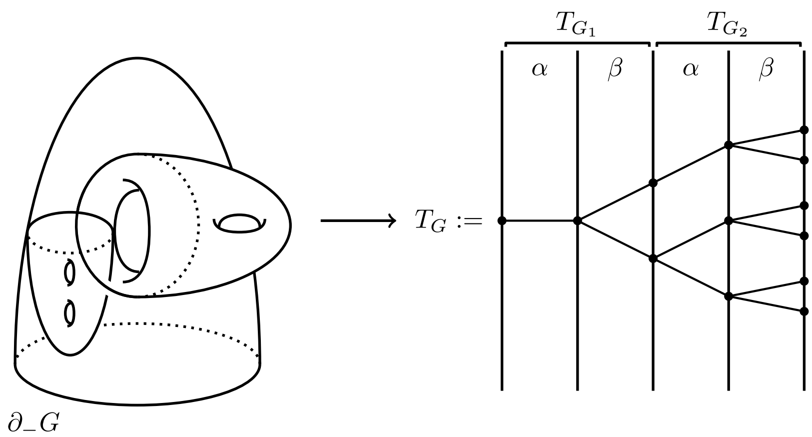


Figure 13.20. A tree associated to a height two grope $G = G_0 \cup G_1 \cup G_2$.

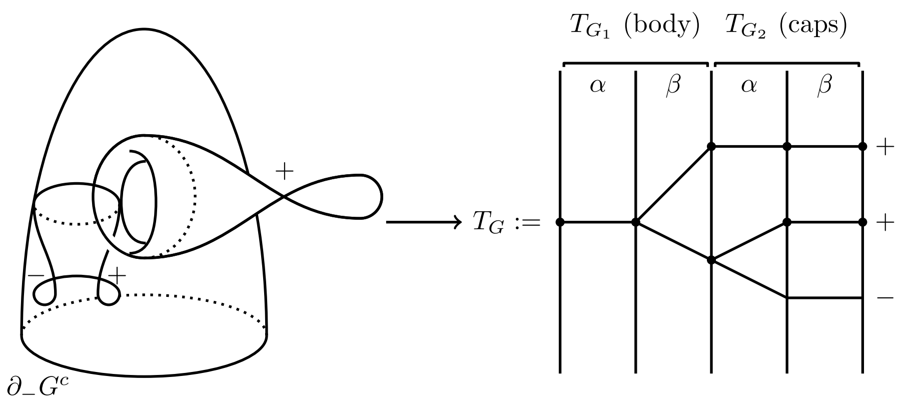


Figure 13.21. A graph associated to a height one capped grope $G^c = G_0 \cup G_1 \cup G_2$.

13.6.4. Kirby diagrams from trees. The recipe for translating a tree to a Kirby diagram for the corresponding generalised tower, along with designated attaching and tip regions, is straightforward. One modifies one of the two components of the Hopf link according to the tree, where the edges in the α regions correspond to ramification, and the edges in the β regions correspond to Bing doubling when coming from a surface block and to Whitehead doubling when coming from a disc or cap block. We explain the iterative process in greater detail next.

Begin with a tree $T_{\mathcal{T}} = T_{G_1} \cup \cdots \cup T_{G_h}$ for a disc-like generalised tower $\mathcal{T} := G_0 \cup G_1 \cup \cdots \cup G_h$ with no additional plumbings across distinct cap blocks. Start with the Hopf link $\widetilde{L} = L \sqcup L^0$. The component L will be fixed throughout the construction of the Kirby diagram, while the component L^0 will be iteratively modified.

Replace L^0 with 0-framed parallel copies of L^0, where the number of copies is equal to the number of edges in the α part of T_{G_1}. Since $\mathcal{T}$ is a disc-like generalised tower, the first stage G_1 is a single block. If it is a surface block, replace the new link components by their Bing doubles. Otherwise, replace them by their Whitehead doubles, with clasp corresponding to the sign of the associated edge in the β part of T_{G_1}. Let L^1 denote the collection of all the new link components produced so far. That is, the original Hopf link $L \sqcup L^0$ has been replaced by the link $L \sqcup L^1$. Note that the link L^1 is an unlink. The Kirby diagram obtained by placing dots on each component of L^1 represents the generalised tower $G_0 \cup G_1$, where the curve L denotes the attaching region. The tip regions consist of 0-framed neighbourhoods of meridians of the components of L^1. By construction, there is a prescribed bijection between the components of L^1 and the tip regions of $G_0 \cup G_1$, since we recorded this information in the tree T_{G_1}.

Next we describe the inductive step. Suppose the subtree $T_{G_1} \cup \cdots \cup T_{G_{i-1}}$ has already been translated into a link $L \sqcup L^{i-1}$. Choose a component ℓ of the link L^{i-1}. This corresponds to a leaf of the tree $T_{G_1} \cup \cdots \cup T_{G_{i-1}}$. Replace ℓ by 0-framed parallel copies of ℓ, where the number of copies is equal to the number of edges in the α part of T_{G_i}. Perform this process for every component of L^{i-1}. From T_{G_i}, we inherit the information of which of these new components arose from surface blocks and which arose from disc or cap blocks. Next, replace each new link component coming from a surface block by its Bing double, and replace each new link component coming from a disc or cap block by its Whitehead double, with the clasp corresponding to the sign of the associated edge in the β part of T_{G_i}. Let L^i denote the collection of the new link components produced in this step. That is, the original link $L \sqcup L^{i-1}$ has been replaced by the link $L \sqcup L^i$. Note that the link L^i is an unlink, since L^{i-1} is. The Kirby diagram obtained by placing dots on each component of L^i represents the generalised tower $G_0 \cup G_1 \cup \cdots \cup G_i$, where the curve L denotes the attaching region. The tip regions consist of 0-framed neighbourhoods of meridians of the components of L^i. By construction, there is a prescribed bijection between the components of L^i and the tip regions of $G_0 \cup G_1 \cup \cdots \cup G_i$, since we recorded this information in the tree $T_{G_1} \cup \cdots \cup T_{G_i}$.

The translation of a tree into a Kirby diagram is illustrated in Figure 13.22. Figure 13.23 shows a Kirby diagram of a capped grope of height two, in other words a 1-storey tower. For a tower with more storeys or caps, replace each dotted circle by its ramified Bing or Whitehead double as dictated by the composition of the tower. We leave this onerous task to the reader, as well as the task of verifying that the Kirby diagrams in this section are obtained from those in Section 13.4 by the simplifications in Section 13.6.

We observe that the links constructed by the process just described are the result of iterated ramified Bing doubling of a single component of the Hopf link when the associated generalised tower is a grope. They are the result of iterated mixed ramified Bing-Whitehead doubling on a single component of the Hopf link when the associated generalised tower is a tower, where the last iteration consists of ramified

Whitehead doubling, corresponding to the final disc stage. Moreover, these are mixed ramified Bing-Whitehead links in a solid torus, formed by the complement of an open tubular neighbourhood of the untouched, unknotted component of the original Hopf link $\widetilde{L} = L \sqcup L^0$. In particular this means these links are of the sort used in the defining sequences for mixed ramified Bing-Whitehead decompositions in Chapter 8.

13.7. Chapter summary

We repeat the key points the reader should recall from this chapter moving forward. First of all, we have constructed Kirby diagrams for towers, consisting only of dotted circles. This implies that every tower or capped tower is diffeomorphic to a boundary connected sum $\natural^m S^1 \times D^3$ for some m. By the interpretation of the dotted circle notation for 1-handles given in Section 13.2.1, towers can be understood explicitly as a subset of D^4. More precisely, let $\{\Delta_i\}$ denote the collection of standard unknotted, disjointly embedded discs in D^4 bounded by the dotted circles in the (simplified) Kirby diagram for the tower. Then the tower is diffeomorphic to $D^4 \smallsetminus \bigcup \Delta_i$, with appropriate corners added for the attaching region. We collect the key points in the following theorem.

THEOREM 13.2. *Let $\mathcal{T}$ be a k-storey tower. A Kirby diagram for $\mathcal{T}$ is given by a mixed ramified Bing-Whitehead link L^k in the solid torus formed by the complement of a 0-framed open tubular neighbourhood of an unknot $U \subseteq S^3$. Each component of L^k is decorated with a dot and the (closed) tubular neighbourhood of the unknot U corresponds to the attaching region $\partial_- \mathcal{T}$. The amount of Bing doubling, Whitehead doubling, and ramification in the diagram correspond to the combinatorics of the tower in the precise manner explained in Section 13.6. In particular, there are k iterations of ramified Bing or Whitehead doubling in the construction of L^k. The surface stages in $\mathcal{T}$ correspond to Bing doubling, and the disc stages in $\mathcal{T}$ correspond to Whitehead doubling. The ramification corresponds to either the genus of the corresponding surface or the number of self-plumbings in the corresponding disc. The signs of the clasps of the Whitehead doubles correspond to the signs of the self-intersections of the disc stages. The last stage in the construction of L^k consists of Whitehead doubling, corresponding to the final disc stage.*

We finish by discussing where to find the tip regions of a tower $\mathcal{T}$ with k storeys in its simplified Kirby diagram as constructed above. Let L^k denote the dotted unlink forming the simplified Kirby diagram for $\mathcal{T}$. Let U denote the curve indicating the attaching region. Let $\{\Delta_i\}$ denote the collection of standard unknotted discs bounded by the components of L^k, so that $\mathcal{T} \cong D^4 \smallsetminus \bigcup \Delta_i$ (modulo corners). So far we have described the tip region as the collection of trivially framed closed tubular neighbourhoods of meridians of the components of L^k. In the next chapter it will be convenient to isotope the tip region slightly. In particular, as we saw in Section 13.2.1, the tip region is isotopic in the boundary $\partial\mathcal{T}$ to the collection of surgery solid tori $\Delta_i \times S^1$. Indeed, the boundary of $\mathcal{T}$ is given by a surgery diagram where we replace all the dots on the components of L^k by zeros.

Then the boundary $\partial\mathcal{T}$ of the tower $\mathcal{T}$ decomposes as the union of three pieces, which one can think of as the foundations, the walls, and the roof of the tower $\mathcal{T}$:

$$\partial\mathcal{T} = \partial_-\mathcal{T} \cup \partial_+\mathcal{T} \cup \overline{\partial}\mathcal{T},$$

where

(i) $\partial_-\mathcal{T} \cong S^1 \times D^2$ is the attaching region, identified with the closed tubular neighbourhood νU;

(ii) $\partial_+\mathcal{T} = \overline{S^3 \smallsetminus (\nu U \sqcup \nu L^k)}$ is the vertical boundary of $\mathcal{T}$; and

(iii) $\overline{\partial}\mathcal{T}$ is the tip region, identified with the union of the collection of closed surgery solid tori $\{\Delta_i \times S^1\}$.

We note that

$$\begin{aligned} \partial_- \mathcal{T} \cap \partial_+ \mathcal{T} &\cong S^1 \times S^1, \\ \partial_- \mathcal{T} \cap \overline{\partial}\mathcal{T} &= \emptyset, \text{ and} \\ \partial_+ \mathcal{T} \cap \overline{\partial}\mathcal{T} &\cong \bigsqcup_i S^1 \times S^1, \end{aligned}$$

where these pairwise disjoint tori are precisely the corners of $\mathcal{T}$. Recall that we introduce corners as described in Section 13.2.3.

So far, we have considered every tower $\mathcal{T}$ as a submanifold of a 0-handle D^4. However, when we introduce corners in $\mathcal{T}$, we also introduce corners in D^4, transforming it to $D^2 \times D^2$. From now on, we refer to towers as being submanifolds of $D^2 \times D^2$, where we assume that the inclusion map respects the corner along the boundary torus of the attaching region.

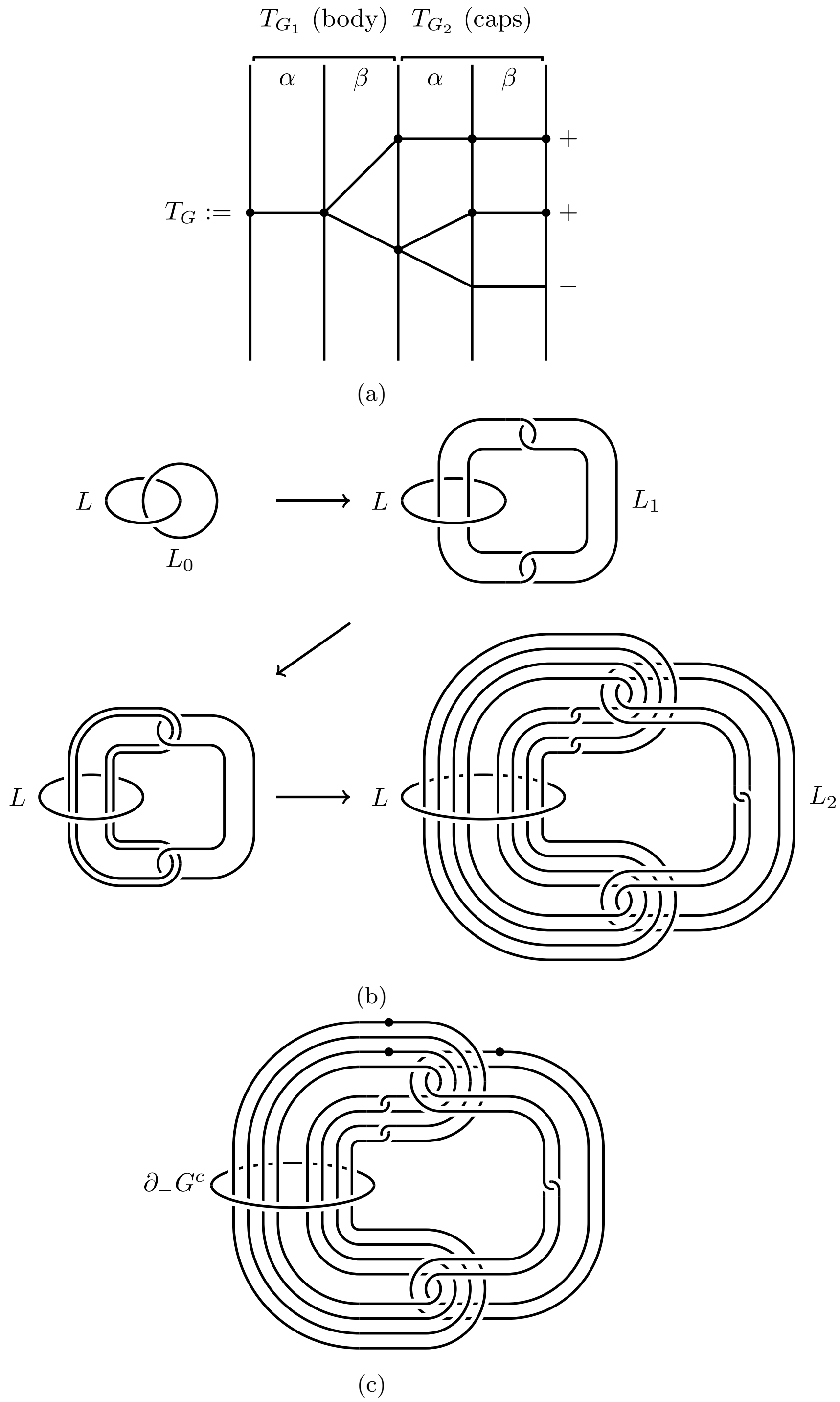


Figure 13.22. Translating a tree into a Kirby diagram. (a) A tree for a height one capped grope G^c. (b) Constructing the corresponding mixed ramified Bing-Whitehead link. (c) The resulting Kirby diagram.

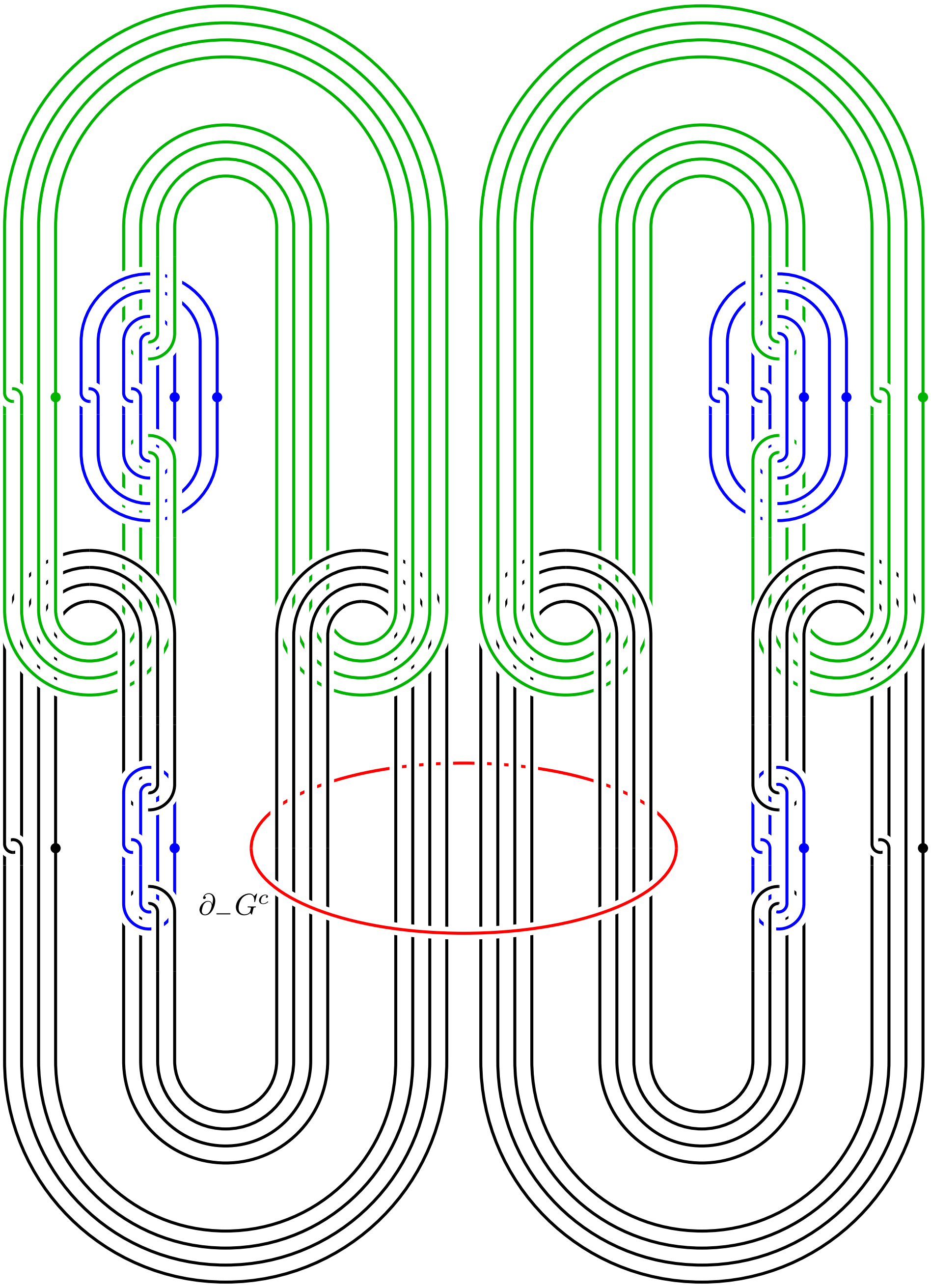


Figure 13.23. A Kirby diagram for a height two capped grope G^c, illustrating some nontrivial ramification. The attaching region $\partial_- G^c$ is a trivially framed neighbourhood of the red circle. We invite the reader to determine the 2-dimensional spine of G^c and to determine where the tip region lies in this diagram.

CHAPTER 14

Architecture of infinite towers and skyscrapers

Stefan Behrens and Mark Powell

We apply the previous chapter to understanding infinite towers. We make the connection between infinite compactified towers and mixed ramified Bing-Whitehead decompositions explicit, and we establish some notation to be used in the remainder of the proof.

14.1. Infinite towers

In Chapter 13, we learnt that a tower $\mathcal{T}$ with k storeys is diffeomorphic, after smoothing corners, to $\natural^m S^1 \times D^3$ for some m (Theorem 13.2). In particular, it has a Kirby diagram given by an m-component dotted link L^k in the complement of an unknot U in S^3. A closed tubular neighbourhood νU is identified with the attaching region of $\mathcal{T}$ and so we may consider L^k to be a link in the solid torus complement of νU. The mixed ramified Bing-Whitehead link L^k is determined up to isotopy by the combinatorics of the tower $\mathcal{T}$, namely the genera of the surfaces in the surface stages and the number and signs of the double points in the disc stages. Considered as a link in S^3 and forgetting U, the link L^k is trivial, and thus the components bound a collection of mutually disjoint, embedded, standard unknotted discs δ_i^k in $D^2 \times D^2$, $i = 1, \dots, m$. We saw that

$$\mathcal{T} \cong D^2 \times D^2 \smallsetminus \big(\coprod_{i=1}^{m} \delta_i^k \times D^2 \big)$$

and that the pairwise disjoint solid tori $\{\delta_i^k \times \partial D^2\}$ together form the tip region of $\mathcal{T}$.

Now consider an infinite tower $\mathcal{T}_\infty$. By definition, an infinite tower is the colimit, or infinite union, of the finite towers obtained as truncations. The Kirby diagrams for these truncations can be related to one other in a precise way. Let $\mathcal{T}_\infty^{\leq k}$ denote the tower consisting of the first k storeys of $\mathcal{T}_\infty$. By the method of construction described in the previous chapter, the mixed ramified Bing-Whitehead link L^{k+1} for $\mathcal{T}_\infty^{\leq k+1}$ is obtained from the mixed ramified Bing-Whitehead link L^k for $\mathcal{T}_\infty^{\leq k}$ by applying some more iterations of ramified Bing doubling, followed by a final iteration of ramified Whitehead doubling. In other words, the components of the link L^{k+1} lie within a tubular neighbourhood of the link L^k. In terms of the Kirby diagram for $\mathcal{T}_\infty^{\leq k+1}$, we may consider filling back in the thickened discs $\{\delta_i^k \times D^2\}$ bounded by the components of L^k, and then removing thinner thickened discs bounded by the components of L^{k+1} within these. This way, we see that $\mathcal{T}_\infty^{\leq k}$ is a subset of $\mathcal{T}_\infty^{\leq k+1}$, respecting the attaching regions and the decomposition into storeys, as desired.

Then we have nested collections of thickened discs $\{\delta_i^k \times D^2\}_{i,k}$, properly embedded in $D^2 \times D^2$, bounded by the sequence $\{L^k\}_k$ of mixed ramified Bing-Whitehead links corresponding to the truncations $\{\mathcal{T}_\infty^{\leq k}\}_k$ of the infinite tower $\mathcal{T}_\infty$.

Recall from Definition 4.3 that a *decomposition* of a topological space is a collection of mutually disjoint subsets, called the decomposition elements. Let $\mathcal{D}_{\mathcal{T}_\infty}$ denote the decomposition of $D^2 \times D^2$ consisting of the connected components of the infinite intersection of $\{\delta_i^k \times D^2\}_{i,k}$. Then

$$\mathcal{T}_\infty \cong D^2 \times D^2 \smallsetminus \bigcup_{\Delta_i \in \mathcal{D}_{\mathcal{T}_\infty}} \Delta_i.$$

In other words, an infinite tower can be understood explicitly as a subset of $D^2 \times D^2$, namely $D^2 \times D^2 \smallsetminus \bigcup \mathcal{D}_{\mathcal{T}_\infty}$, respecting the corner at the boundary of the attaching region.

By this description, the boundary $\partial\mathcal{T}_\infty$ is diffeomorphic, after smoothing corners, to the complement in S^3 of the union of the elements of a mixed ramified Bing-Whitehead decomposition, from Chapter 8, as follows. The boundary of $\mathcal{T}_\infty$ splits into two pieces

$$\partial\mathcal{T}_\infty = \partial_-\mathcal{T}_\infty \cup_{S^1 \times S^1} \partial_+\mathcal{T}_\infty,$$

where $\partial_-\mathcal{T}_\infty \cong S^1 \times D^2$ is the attaching region and $\partial_+\mathcal{T}_\infty$ is the vertical boundary. From the description of $\mathcal{T}_\infty$, we also see that the attaching region $\partial_-\mathcal{T}_\infty$ is the (closed) tubular neighbourhood νU. We had tubular neighbourhoods νL^k of the mixed ramified Bing-Whitehead links $\{L^k\}_k$ such that νL^{k+1} is contained in νL^k for each k. Let $\partial\mathcal{D}_{\mathcal{T}_\infty}$ denote the mixed ramified Bing-Whitehead decomposition of S^3 consisting of the connected components of the infinite intersection of $\{\nu L^k\}_k$. Then

$$\partial\mathcal{T}_\infty \cong S^3 \smallsetminus \bigcup_{\Delta_i \in \partial\mathcal{D}_{\mathcal{T}_\infty}} \Delta_i.$$

Moreover the vertical boundary $\partial_+\mathcal{T}_\infty$ is the complement of $\bigcup_{\Delta_i \in \partial\mathcal{D}_{\mathcal{T}_\infty}} \Delta_i$ in the solid torus $\overline{S^3 \smallsetminus \nu U}$.

14.2. Infinite compactified towers

PROPOSITION 14.1. *The endpoint compactification $\widehat{\mathcal{T}}_\infty$ of an infinite tower $\mathcal{T}_\infty$ is homeomorphic to the decomposition space $D^2 \times D^2/\mathcal{D}_{\mathcal{T}_\infty}$ where $\mathcal{D}_{\mathcal{T}_\infty}$ denotes the decomposition of $D^2 \times D^2$ consisting of the connected components of the infinite intersection of $\{\delta_i^k \times D^2\}_{i,k}$ defined in the previous section.*

Here we use the convention of Part I (see Definition 4.3) that the elements of $\mathcal{D}_{\mathcal{T}_\infty}$ are crushed to distinct individual points (rather than all being crushed to the same point) in the decomposition space $D^2 \times D^2/\mathcal{D}_{\mathcal{T}_\infty}$.

Note that we do not conclude that the infinite compactified tower $\widehat{\mathcal{T}}_\infty$ is diffeomorphic to the decomposition space $D^2 \times D^2/\mathcal{D}_{\mathcal{T}_\infty}$. In particular, *a priori* neither space is known to have a smooth structure, or even to be manifold at all.

PROOF. By Definition 12.17, the ends of $\mathcal{T}_\infty$ correspond to sequences of connected components of complements of finite truncations of $\mathcal{T}_\infty$, which correspond to paths in the associated graph starting at the root, which in turn correspond to elements of the decomposition $\mathcal{D}_{\mathcal{T}_\infty}$. In other words, the diffeomorphism

$$D^2 \times D^2 \smallsetminus \bigcup_{\Delta_i \in \mathcal{D}_{\mathcal{T}_\infty}} \Delta_i \cong \mathcal{T}_\infty$$

extends to a function $f\colon D^2 \times D^2 \to \widehat{\mathcal{T}}_\infty$ by sending each element of $\mathcal{D}_{\mathcal{T}_\infty}$ to the corresponding endpoint of $\widehat{\mathcal{T}}_\infty$.

We claim that f is continuous. This follows from the definition of $\widehat{\mathcal{T}}_\infty$. We need to check that the preimage of every open set in $\widehat{\mathcal{T}}_\infty$ is open in $D^2 \times D^2$. This is obviously true for every open set in $\mathcal{T}_\infty \subseteq \widehat{\mathcal{T}}_\infty$. By the definition of endpoint compactification, it remains to check the open sets $V \subseteq \widehat{\mathcal{T}}_\infty$ such that $V \cap \mathcal{T}_\infty = U_j$ for some j and $V \cap (\widehat{\mathcal{T}}_\infty \smallsetminus \mathcal{T}_\infty)$ consists of sequences $(V_1, V_2, \dots)$ such that $V_{j+i} \subseteq U_j$ for all $i > 0$, where U_j and each V_k is a connected component of the complement of the truncation $\mathcal{T}_\infty^{\leq j}$ and $\mathcal{T}_\infty^{\leq k}$ respectively. The connected components of the complements in $D^2 \times D^2$ of finite truncations of $\mathcal{T}_\infty$ are precisely (open) thickened discs in $D^2 \times D^2$. These are open neighbourhoods of the elements of $\mathcal{D}_{\mathcal{T}_\infty}$, completing the proof that f is continuous. Then we have a diagram

$$\begin{tikzcd} D^2 \times D^2 \arrow[r, "f"] \arrow[d, "\pi"] & \widehat{\mathcal{T}}_\infty \\ D^2 \times D^2/\mathcal{D}_{\mathcal{T}_\infty} \arrow[ur, "\overline{f}"'] & \end{tikzcd}$$

where the vertical map π is the quotient and $\overline{f}$ is a continuous bijection since f is constant on the fibres of π. We saw in Section 12.2 that $\widehat{\mathcal{T}}_\infty$ is Hausdorff. Since $D^2 \times D^2$ is compact, so is the quotient $D^2 \times D^2/\mathcal{D}_{\mathcal{T}_\infty}$. Since $\overline{f}$ is a continuous bijection from a compact space to a Hausdorff space, it is a homeomorphism by the closed map lemma (Lemma 3.23). $\square$

Proposition 14.1 also allows us to see the boundary of an infinite compactified tower in terms of decomposition spaces. Recall that we have

$$\partial\widehat{\mathcal{T}}_\infty = \partial_-\widehat{\mathcal{T}}_\infty \cup_{S^1 \times S^1} \partial_+\widehat{\mathcal{T}}_\infty,$$

where $\partial_-\widehat{\mathcal{T}}_\infty$ is the attaching region and $\partial_+\widehat{\mathcal{T}}_\infty$ is the vertical boundary. By definition, the attaching region of an infinite compactified tower $\widehat{\mathcal{T}}_\infty$ is the attaching region of the underlying infinite tower $\mathcal{T}_\infty$, and is thus identified with a closed tubular neighbourhood νU of an unknot U in S^3. The vertical boundary $\partial_+\widehat{\mathcal{T}}_\infty$ is by definition (Remark 12.19) the endpoint compactification of the vertical boundary $\partial_+\mathcal{T}_\infty$ of $\mathcal{T}_\infty$.

By a virtually identical proof as for Proposition 14.1, the boundary $\partial\widehat{\mathcal{T}}_\infty$ is homeomorphic to the decomposition space $S^3/\partial\mathcal{D}_{\mathcal{T}_\infty}$ where $\partial\mathcal{D}_{\mathcal{T}_\infty}$ denotes the mixed ramified Bing-Whitehead decomposition of S^3 consisting of the connected components of the infinite intersection of $\{\nu L^k\}_k$ defined in the previous section. The vertical boundary $\partial_+\widehat{\mathcal{T}}_\infty$ is the decomposition space $\overline{S^3 \smallsetminus \nu U}/\partial\mathcal{D}_{\mathcal{T}_\infty}$.

14.3. Skyscrapers

Recall that by definition a skyscraper is a particular type of infinite compactified tower (Definition 12.21). As a result all of our work in the previous section applies to them. In particular, a skyscraper is a disc-like infinite compactified tower satisfying the following two conditions.

(a) (Replicable) Each storey has at least four surface stages.
(b) (Boundary shrinkable) the series $\sum_{i=1}^{\infty} N_j/2^j$ diverges, where N_j is the number of surface stages in the jth storey.

The boundary shrinkable condition implies, via Theorem 8.1, that the corresponding mixed ramified Bing-Whitehead decomposition of S^3 shrinks, and so for any skyscraper $\widehat{\mathcal{S}}$ we have a homeomorphism

$$\partial\widehat{\mathcal{S}} \cong S^3/\partial\mathcal{D}_{\mathcal{S}} \cong S^3,$$

where $\partial\mathcal{D}_{\mathcal{S}}$ denotes the mixed ramified Bing-Whitehead decomposition of S^3 consisting of the connected components of the infinite intersection of $\{\nu L^k\}_k$, defined in Section 14.1, corresponding to the open skyscraper $\mathcal{S}$. In particular $\partial\widehat{\mathcal{S}}$ is a manifold. Moreover, the vertical boundary is

$$\partial_+\widehat{\mathcal{S}} = \overline{\partial\widehat{\mathcal{S}} \smallsetminus \partial_-\widehat{\mathcal{S}}} \cong D^2 \times S^1.$$

Now let $\widehat{\mathcal{S}}$ be a skyscraper. Recall that $\widehat{\mathcal{S}}^{\leq k}$ denotes the finite tower obtained as a truncation of $\widehat{\mathcal{S}}$ consisting of the union of the first k levels, that is the first $2k$ storeys when $\widehat{\mathcal{S}}$ is considered as a compactified infinite tower. Also recall that $\partial_+\widehat{\mathcal{S}}^{\leq k}$ denotes the vertical boundary, which is the closure of the complement in the boundary $\partial\widehat{\mathcal{S}}^{\leq k}$ of the attaching region $\partial_-\widehat{\mathcal{S}}^{\leq k}$ and the tip regions $\overline{\partial}\widehat{\mathcal{S}}^{\leq k}$:

$$\partial_+\widehat{\mathcal{S}}^{\leq k} = \overline{\partial\widehat{\mathcal{S}}^{\leq k} \smallsetminus (\partial_-\widehat{\mathcal{S}}^{\leq k} \cup \overline{\partial}\widehat{\mathcal{S}}^{\leq k})}.$$

PROPOSITION 14.2. *Each $\partial_+\widehat{\mathcal{S}}^{\leq k}$ is diffeomorphic to the complement of an open tubular neighbourhood of a mixed ramified Bing-Whitehead link in $D^2 \times S^1$ with Whitehead doubles in its last stage.*

To obtain $\partial_+\widehat{\mathcal{S}}^{\leq k+1}$ from $\partial_+\widehat{\mathcal{S}}^{\leq k}$, add in the neighbourhood of the mixed ramified Bing-Whitehead link, and remove the next link in the nested sequence instead.

CHAPTER 15

Basic geometric constructions

Mark Powell and Arunima Ray

We have now described the finite and infinite iterated objects that will appear in the proof of the disc embedding theorem. In this chapter, we gather some tools that we will use to build them.

Throughout this chapter we work in an ambient smooth 4-manifold M.

15.1. Clifford torus

In a neighbourhood of a double point of immersed surfaces A and B (where possibly $A = B$) in M, lies an embedded torus, as shown in Figure 15.1. We call this torus the *Clifford torus* of the double point. More concretely, the local picture at the double point is modelled by the xy- and zt-planes intersecting at the origin in $\mathbb{R}^4$, and the Clifford torus corresponds to $S^1 \times S^1$, where the first S^1 factor is the unit circle in the xy-plane and the second is the unit circle in the zt-plane. The fundamental group of this torus is generated by meridians of the two surfaces. Clifford tori are equipped with a canonical framing by construction.

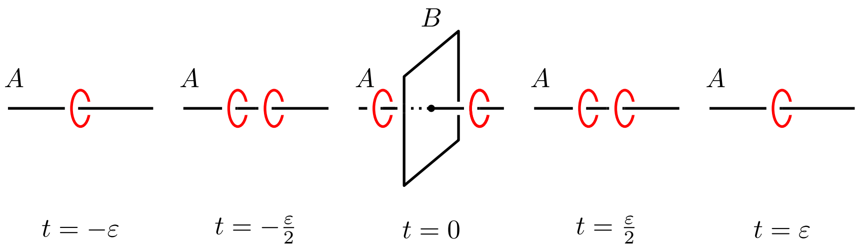


Figure 15.1. A Clifford torus at a double point of the surfaces A and B. As usual the fourth coordinate is interpreted as time t. The double point is seen at the $t = 0$ time slice. The red circles depicted trace out the Clifford torus as we move backwards and forwards in time. Note that each S^1 factor of the torus corresponds to a meridian of either A or B, and as such bounds a meridional disc that intersects the respective surface transversely at a single point. Compare with Figure 1.5.

A finger move from an immersed surface A to an immersed surface B within the ambient 4-manifold M, where possibly $A = B$, changes the fundamental group of $M \setminus (A \cup B)$ by adding the relation corresponding to the 2-cell of the Clifford torus of either of the newly introduced double points. The standard argument for this [**Cas86**] goes as follows. Let A' denote the surface A after the finger move. After the finger move, there is an embedded, framed Whitney disc D for the two new intersection points between A' and B, the Whitney move along which would undo the finger move. Observe that $\pi_1(M \smallsetminus (A \cup B)) \cong \pi_1(M \smallsetminus (A' \cup B \cup D))$. The Whitney disc D intersects the Clifford torus T in a single point, so there is a meridional disc of D that is also a disc in T. The boundary of this disc is a meridian for D and is freely homotopic in $M \smallsetminus (A' \cup B \cup D)$ to the attaching circle for the 2-cell of T. This attaching circle corresponds to a commutator $[\mu_A, \mu_B]$, where μ_A and μ_B are appropriately based meridians of A and B respectively. Let μ_D be a (based) meridian of D.

Given a group G and $g \in G$, let $\langle\langle g \rangle\rangle$ denote the normal subgroup of G generated by g. We have that:

$$\begin{aligned}\pi_1(M \smallsetminus (A' \cup B)) &\cong \pi_1(M \smallsetminus (A' \cup B \cup D))/\langle\langle \mu_D \rangle\rangle \\ &\cong \pi_1(M \smallsetminus (A' \cup B \cup D))/\langle\langle \gamma_T \rangle\rangle \\ &\cong \pi_1(M \smallsetminus (A \cup B))/\langle\langle [\mu_A, \mu_B] \rangle\rangle,\end{aligned}$$

as asserted.

15.2. Elementary geometric techniques

The following operations will be described on (immersed) surfaces. It will be shown, where necessary, that these operations can be extended to framed surfaces, and therefore to stages of capped or uncapped gropes or towers. However we will not comment on this every time, unless there is extra particular care that needs to be taken.

15.2.1. Tubing. Tubing was described in [**FQ90**, Sections 1.8 and 1.9]. Suppose we have an immersed connected surface A. Let Σ and Σ' be two other immersed surfaces intersecting A transversely at points p and p' respectively with $p \neq p'$. Let γ be a smooth, embedded arc in A joining p and p'. Consider a tubular neighbourhood of γ intersecting A and B in small discs about p and p'. Cut out these discs from $A \cup B$ and glue on the rest of the boundary of the neighbourhood of γ to $\Sigma \cup \Sigma'$. In other words, we are gluing in a meridional annulus for γ. This process is called *tubing* Σ *into* Σ' (*along* $\gamma \subseteq A$) or vice versa. (In [**FQ90**], the resulting surface is called the *sum* of Σ and Σ'.) We shall at times also refer to this process as tubing p into the surface Σ'. If Σ and Σ' are distinct framed immersed surfaces, then the result of tubing is itself immersed and inherits a framing coming from the framing of the annulus, of Σ, and of Σ'. If Σ and Σ' are the same immersed framed surface we need additionally that the intersection points p and p' have opposite signs, otherwise we would obtain a nonorientable surface. Note that the last sentence is still meaningful even if the ambient space M is nonorientable, by using the arc γ to transport a local orientation at p to one at p'.

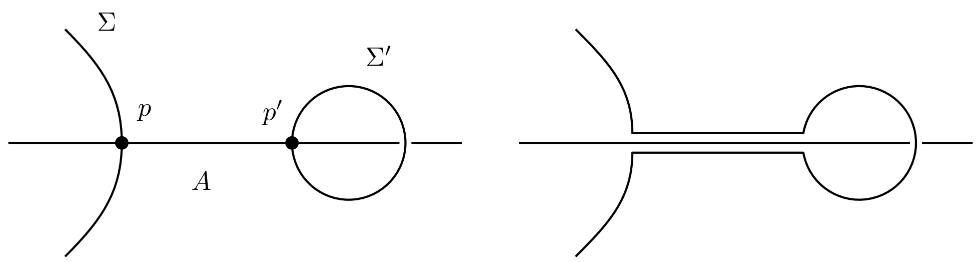


Figure 15.2. Tubing into a transverse sphere.

We will often tube surfaces into geometrically transverse spheres, that is, Σ' will be a framed sphere such that p' is its unique point of intersection with A. This situation is depicted in Figure 15.2. Note that this operation then decreases the number of points of intersection between A and Σ by one.

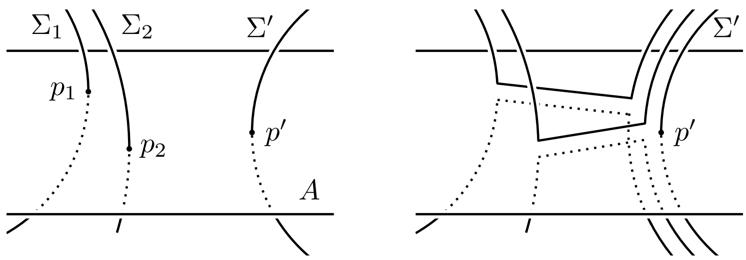


Figure 15.3. Tubing multiple points of intersection.

In most cases, we will tube into a parallel push-off of Σ' instead of Σ' itself. This will allow us to tube multiple times, as follows. Suppose that Σ' is a framed surface geometrically transverse to A, that is A and Σ' intersect exactly once, at the point p'. Now suppose we have surfaces Σ_i intersecting A at the points p_i, for $i = 1, \ldots, k$. Then take k parallel push-offs of Σ', using distinct nonvanishing sections of the normal bundle (of the form $\Sigma' \times \{p\}$ in the given framing). These intersect A at k distinct points. Find pairwise disjoint embedded arcs on A joining these points to the $\{p_i\}$ and use these to tube each Σ_i into a distinct parallel push-off of Σ'. This eliminates the intersection points $p_1, \ldots, p_k$, and also leaves the original Σ' unchanged for further tubing or other constructions. This process is described in Figure 15.3.

Note that if some surface B intersects Σ', then now Σ_i intersects B for each i. In particular, if Σ' has double points, then the tubing creates new points in $\Sigma_i \pitchfork \Sigma_j$, for all i, j, as well as in $\Sigma_i \pitchfork \Sigma'$.

15.2.2. Boundary twisting. Boundary twisting was introduced in [**FQ90**, Section 1.3]. Suppose we have two immersed surfaces A and B in a 4-manifold M such that part of the boundary of B is embedded in A, as shown in Figure 15.4(a), and this part of ∂B lies in the interior of M. For us, this situation usually arises when B is a Whitney disc pairing intersection points of A with itself or some other surface. Another commonly occurring situation is that of a cap of a grope or tower attached to a lower stage. The operation of *boundary twisting* B about A consists of changing a collar of B near a point in its boundary on A, as depicted in Figure 15.4(b). Note that this creates a new point of intersection between A and B and changes the framing of B by a full twist.

15.2.3. Making Whitney circles disjoint. We saw in Chapter 11 that given two immersed surfaces A and B, if $\lambda(A, B) = 0$, the intersection points between A and B can be paired up with Whitney discs (Proposition 11.10). *A priori* the corresponding Whitney circles may intersect one another. Performing the Whitney trick with intersecting Whitney circles leads to new intersections, which we would like to avoid. We ensure that Whitney circles are disjoint by pushing one Whitney circle along the other, as shown in Figure 15.5. This is a regular homotopy, and as we see in the figure, leads to new intersections between a Whitney disc and either A or B.

To achieve this in general for surfaces A and B, enumerate the Whitney arcs, and then work on the arcs in order. For the ith arc, push other arcs with index greater

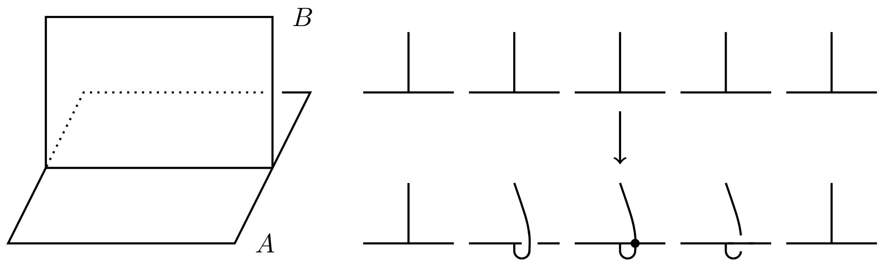


Figure 15.4. Boundary twisting. Left: Before boundary twisting. Right: Cross sections for the picture on the left, before and after boundary twisting. Each cross section shows a 3-dimensional time slice.

than i off the ith arc, starting with one of the arcs closest to the endpoint, until the ith arc is disjoint from all other arcs. At the end of the process, all Whitney arcs are mutually disjoint. From now on, we always assume that Whitney circles are disjoint without comment or loss of generality.

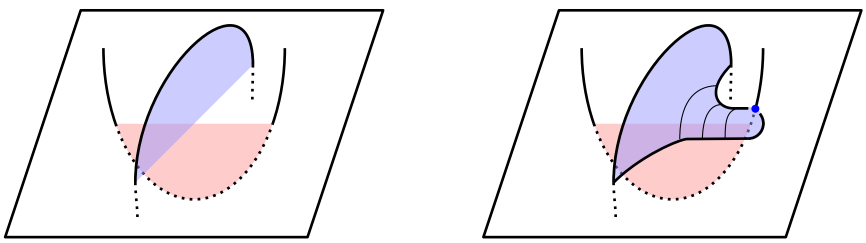


Figure 15.5. Boundary push off to ensure Whitney circles are disjoint. We perform a regular homotopy of one Whitney circle until it becomes disjoint from the other, introducing a new intersection point for the corresponding Whitney disc.

Remark 15.1. The last two techniques show how to complete the proof of Proposition 11.10. In that proposition, we obtained immersed Whitney discs using algebraic topological considerations, due to the vanishing of the appropriate intersection and self-intersection numbers. These discs come from null homotopies and thus they may not *a priori* be framed. The boundary twisting operation allows us to ensure that the Whitney discs are framed, at the expense of adding new points of intersection between the Whitney discs and the original surfaces. Since such intersections are often already present, or at least cannot be assumed not to be present, this does not hurt us in practice. The boundary of the Whitney discs may not be embedded or mutually disjoint to begin with but this can be ensured by the procedure of the previous section.

15.2.4. Pushing down intersections. The technique of pushing down intersection points was introduced in [**FQ90**, Section 2.5]. Suppose we have two immersed surfaces A and B such that part of the boundary of B is embedded in A, as shown in Figure 15.6(a) (note this is the same situation as in Figure 15.4).

Then any intersection between B and some third surface C can be removed at the expense of adding two new intersections between C and A. This is shown in Figure 15.6(b). For us, most often A will be part of a surface stage in a capped grope or tower, and B will be part of either a cap stage or a surface stage. We can then iteratively push down intersections with B to any surface stage including or below A. One advantage of doing this is that the new intersections appear in algebraically cancelling pairs and thus have associated Whitney discs. Alternatively, often a lower stage of a grope will have a geometrically transverse sphere, and so we can push down the intersection points and then tube the many new intersection points into the geometrically transverse sphere.

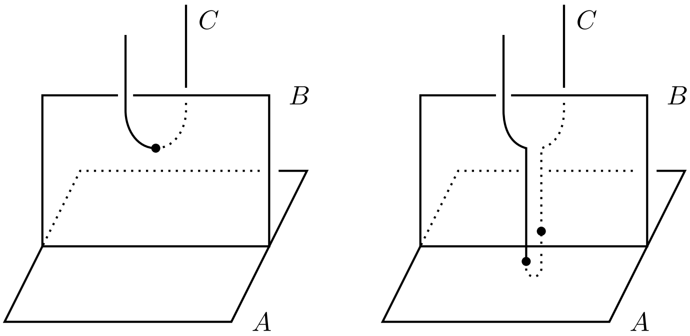


Figure 15.6. Pushing down an intersection point between C and B (left) leads to two new intersection points between C and A (right). Note that the two new intersection points are evidently paired by an embedded, framed Whitney disc. Only a single time slice is pictured. Note that we are performing a finger move, albeit one that pushes across the boundary and therefore does not preserve intersection numbers.

15.2.5. Contraction and subsequent pushing off. Contraction and push off was introduced in [**FQ90**, Section 2.3]. The (symmetric) contraction of a capped surface, depicted in Figure 15.7, converts a capped surface into an immersed disc. As shown by the figure, we start with a symplectic basis of curves on the surface and surger the surface using two copies each of framed immersed discs bounded by these curves joined by a square at the point of intersection of the curves. One could alternatively contract a capped surface by only surgering along one disc per dual pair, but this would not enable the pushing off procedure that we are about to describe in the next paragraph. Henceforth, whenever we talk about contraction, by default we will mean the symmetric contraction. Observe that the result of contracting a capped surface with embedded body has algebraically cancelling self-intersections.

After contracting a capped surface Σ^c with body Σ, any other surface A that intersected the caps of Σ^c can be pushed off the contracted surface, as we describe in Figure 15.8. The fact that we can perform the pushing off procedure, which is a regular homotopy, shows that the intersection number of the contracted surface with A is trivial. The push off procedure reduces the number of intersection points between the contracted surface (an immersed disc) and the pushed off surfaces, so we gain some disjointness at the expense of converting a capped surface into an immersed disc. An additional cost is as follows. Suppose that a surface A intersects a cap of the capped surface, and a surface B intersects a dual cap. Then after pushing both A and B off the contraction, we obtain two intersection points between A and B. The contraction push off operation is shown, via before and after pictures, in Figure 15.8.

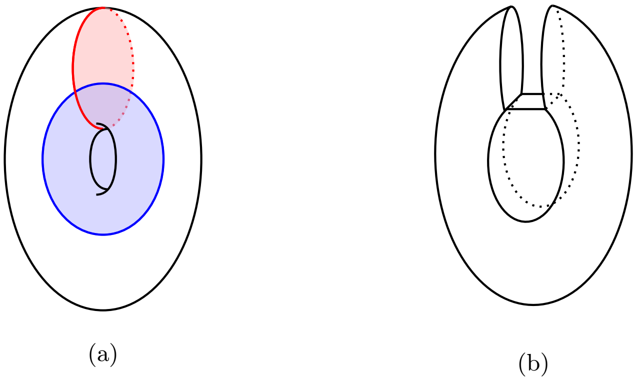


Figure 15.7. (Symmetric) contraction of a capped surface. Here we show the situation for embedded caps.

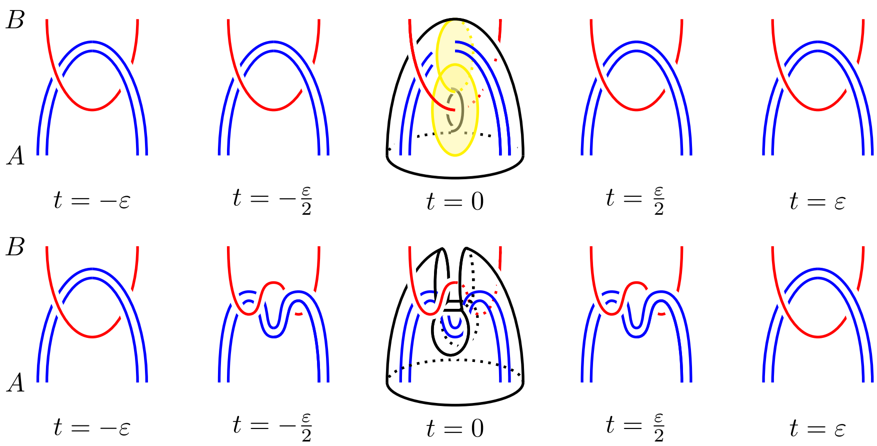


Figure 15.8. Top: Before contraction of a surface. Bottom: After contraction, with other surfaces pushed off the result of contraction. The capped surface being contracted is shown in the middle time slice. The surfaces A and B being pushed off the contraction are shown in blue and red respectively. Note that the intersections of the pushed off surfaces occur between diagrams one and two and between diagrams four and five in the bottom row of figures, namely one intersection in the past and one intersection in the future between each pair of surfaces that were pushed off dual caps.

A useful observation is that the homotopy class of the surface resulting from a contraction of a capped surface is independent of the choice of caps.

Lemma 15.2. *The homotopy class of the sphere or disc resulting from symmetric contraction of a fixed surface is independent of the choice of caps, provided the boundaries of the different choices of caps coincide.*

PROOF. As explained in [**FQ90**, Section 2.3], an isotopy in the model capped surface induces a homotopy of the immersed models. In the model, the symmetric contraction along dual caps $\{C, D\}$ is isotopic to the result of surgery along either cap, for example C. This can be seen directly by isotoping across the region lying between the parallel copies of D used in the symmetric contraction (see Figure 15.7). Therefore once the model is immersed in a 4-manifold, the symmetric contraction is homotopic to the result of surgery along one cap per dual pair.

Now let $\{C_i, D_i\}_{i=1}^g$ and $\{C_i', D_i'\}_{i=1}^g$ be two sets of caps for a surface of genus g, such that $\partial C_i = \partial C_i'$ and $\partial D_i = \partial D_i'$ form a dual pair of curves on the surface for each i. Then the result of contraction along $\{C_i, D_i\}_{i=1}^g$ is homotopic to the asymmetric contraction along $\{C_i\}$, which is homotopic to contraction along $\{C_i, D_i'\}_{i=1}^g$. This is in turn homotopic to the result of asymmetric contraction along $\{D_i'\}$, which finally is homotopic to the result of contraction on $\{C_i', D_i'\}_{i=1}^g$, as asserted. □

15.3. Replacing algebraic duals with geometric duals

We finish the chapter with two applications of the techniques introduced so far. The first lemma will be used repeatedly in the upcoming constructions. It allows us to improve algebraic duals into geometric duals, at the cost of introducing new self-intersections, and first appeared in this form in [**Fre82a**, Lemma 3.1]. Compare with the techniques described in Chapter 1 and see also [**FQ90**, Section 1.5].

LEMMA 15.3 (Geometric Casson lemma). *Let M be a smooth 4-manifold. Let $\{f_i\}$ and $\{g_i\}$ be immersed finite collections of discs or spheres in M, transversely intersecting in their interiors in double points, with $\lambda(f_i, g_j) = \delta_{ij}$ for all i, j. Then there exist families $\{f_i'\}$ and $\{g_i'\}$ of immersed discs or spheres in M, again transversely intersecting only in their interiors in double points, such that:*

- *(i) for every i, f_i' is regularly homotopic to f_i;*
- *(ii) for every i, g_i' is regularly homotopic to g_i;*
- *(iii) the surfaces f_i' and g_i' intersect exactly once transversely; and*
- *(iv) the surfaces f_i' and g_j' are disjoint whenever $i \neq j$.*

Similarly, given distinct families $\{f_i\}$ and $\{g_i\}$ of immersed discs or spheres such that $\lambda(f_i, g_j) = 0$ for all i, j, there exist pairwise disjoint families $\{f_i'\}$ and $\{g_i'\}$ such that for each i, f_i' and g_i' are regularly homotopic to f_i and g_i respectively.

Note that since the lemma provides a regular homotopy, $\lambda(f_i, h) = \lambda(f_i', h)$, $\mu(f_i) = \mu(f_i')$, $\lambda(g_i, h) = \lambda(g_i', h)$, and $\mu(g_i) = \mu(g_i')$ for every i and for every immersed disc or sphere h in M.

PROOF. Since $\lambda(f_i, g_i) = 1$ for each i, we can pair up all but one of the points of intersection with framed, immersed Whitney discs (Proposition 11.10). Similarly, since $\lambda(f_i, g_j) = 0$ whenever $i \neq j$, all of the points of intersection between each f_i and each g_j are paired by Whitney discs. Note that each Whitney circle for the Whitney discs mentioned above has an arc lying in $\{f_i\}$ and an arc lying in $\{g_i\}$.

Push all the points of intersection between the Whitney discs and $\{f_i\}$ and $\{g_i\}$ off the Whitney discs. Do this so that the only new intersections are within each family. That is, push each intersection of a Whitney disc with an element of $\{f_i\}$ onto $\{f_i\}$ by pushing towards the arc of the Whitney circle lying in $\{f_i\}$. Do the same for the $\{g_i\}$, as shown in Figure 15.9. Now we have framed, immersed Whitney discs with interiors lying in the complement of $\{f_i\} \cup \{g_i\}$, and we use them to perform the Whitney move. For each Whitney move push f_i over the Whitney disc. This introduces new intersections among the $\{f_i\}$ coming from the intersection among the Whitney discs, but this is the price we agreed to pay. Note that surfaces or

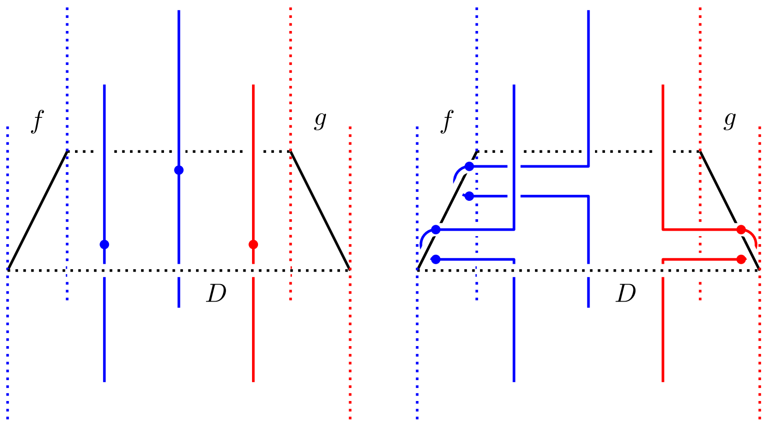


Figure 15.9. Proof of the geometric Casson lemma. Left: The Whitney disc D pairs two intersection points between the surfaces f (blue) and g (red). Only a small portion of the Whitney disc is shown. Right: Push off f intersections into f, and g intersections into g.

curves disjoint from the Whitney discs may be assumed to be unaffected by the construction, by doing everything in a small enough neighbourhood of the Whitney discs. By Proposition 11.1, proved next, the immersed Whitney move is a regular homotopy.

A virtually identical proof gives the second statement. □

More generally, the above argument can be used to ensure that the algebraic and geometric intersection numbers between finite collections $\{f_i\}$ and $\{g_i\}$ agree, at the expense of increasing the number of geometric intersections within each family. On the other hand, we certainly may *not* use this argument to realise a self-intersection number $\mu(f) = 0$ geometrically.

Finally we give a proof of Proposition 11.1, which we now recall.

Proposition 11.1. *A (framed) immersed Whitney move is a regular homotopy.*

Proof. By the definition of regular homotopy, it suffices to show that a framed immersed Whitney move is a sequence of isotopies, finger moves, and Whitney moves along framed, embedded discs with interiors disjoint from the surfaces being homotoped. This is shown in Figure 15.10. First push down self-intersections of W to A (or B) along the collar of W to produce a framed, embedded Whitney disc, whose interior might still intersect $A \cup B$. Observe this does not change A or B. Then, as in the proof of the geometric Casson lemma, push intersections of A with W towards A, and those of B with W along B. Observe that these are finger moves on A and B respectively. The result is a framed, embedded Whitney disc with interiors disjoint from $A \cup B$. Perform the Whitney move on A along this disc. Finally reverse the effect of the previous finger moves on A and B to see that the result coincides with the result of the immersed Whitney move on the original Whitney disc and A and B. More precisely, this final move consists of an isotopy on A and Whitney moves on B along framed, embedded Whitney discs with interiors disjoint from B. □

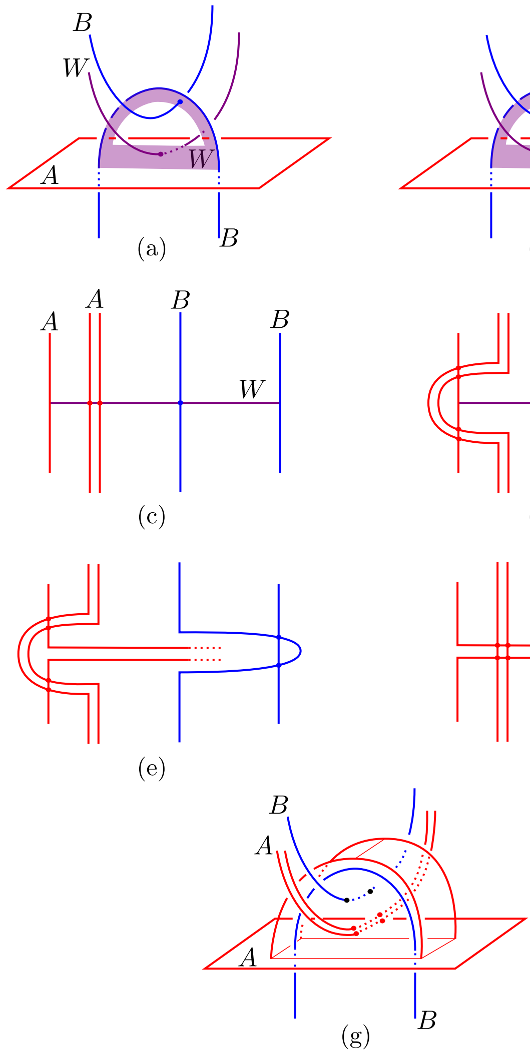


Figure 15.10. (a) A framed, immersed Whitney disc W (purple) pairing intersections between surfaces A (red) and B (blue). The disc W is not embedded, but an embedded collar of its boundary is shown. (b)-(f) A sequence of isotopies, finger moves, and Whitney moves along framed, embedded Whitney discs with interiors disjoint from $A \cup B$. (g) The result of the immersed Whitney move on (a).

CHAPTER 16

From immersed discs to capped gropes

Wojciech Politarczyk, Mark Powell, and Arunima Ray

Now we embark on the proof of the disc embedding theorem in earnest. Recall that we wish to construct embedded skyscrapers in the ambient 4-manifold, given the hypotheses of the disc embedding theorem, such that once we have proven that skyscrapers are homeomorphic to standard 2-handles, we will be able to use these 2-handles to perform Whitney moves. We break down the proof of this into a sequence of propositions, the first two of which appear in this chapter. First we show in Proposition 16.1 how to replace the immersed discs in the hypotheses of the disc embedding theorem by new ones (with the same framed boundary) whose intersections and self-intersections are paired by framed, immersed Whitney discs equipped with geometrically transverse capped surfaces. These Whitney discs are improved to capped gropes (Proposition 16.2), then to a union-of-discs-like capped tower (Proposition 17.12), and finally to mutually disjoint skyscrapers (Proposition 18.12). The full proof for the disc embedding theorem, modulo our work in Part IV is given in Section 18.4.

We begin with the following proposition.

Proposition 16.1. *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 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 an immersed collection

$$F' = (f'_1, \dots, f'_n)\colon (D^2 \sqcup \cdots \sqcup D^2, S^1 \sqcup \cdots \sqcup S^1) \looparrowright (M, \partial M)$$

of discs in M, such that the following holds.

1. *For each i, the discs f_i and f'_i have the same framed boundary.*
2. *There is an immersed collection*
$$G' = (g'_1, \dots, g'_n)\colon S^2 \sqcup \cdots \sqcup S^2 \looparrowright M$$
of framed geometrically transverse spheres for F' in M.
3. *The spheres g'_i and g_i are regularly homotopic for each i.*
4. *$\pi_1(M \smallsetminus F') \cong \pi_1(M)$, and is thus good.*
5. *The intersections and self-intersections of F' are paired by an immersed collection of Whitney discs $\{W_k\}$.*
6. *For each k, the interior of the Whitney disc W_k lies in $M \smallsetminus F'$.*
7. *There exists a collection $\{\Sigma_k^c\}$ of geometrically transverse capped surfaces for the $\{W_k\}$.*
8. *The sets $\{W_k\}$ and $\{\Sigma_k^c\}$ have the same cardinality.*

(9) For each k, $\Sigma_k^c \subset M \smallsetminus F'$.
(10) For each k, the surface Σ_k, namely the body of Σ_k^c, is contained in a regular neighbourhood of F'.
(11) $\lambda(C_\ell, C_m) = \mu(C_\ell) = 0$ for every pair of caps C_ℓ and C_m of the $\{\Sigma_k^c\}$.

We summarise the statement above schematically in Figure 16.1. Note that the caps of the geometrically transverse capped surfaces miss the Whitney discs $\{W_k\}$, but may intersect one another. The bodies of the $\{\Sigma_k^c\}$ are disjointly embedded.

Here and throughout this and the following two chapters, we attempt to depict all possible types of intersections in our figures; in the interest of clarity (and sanity) we occasionally suppress the proliferation of genus and intersections.

Note that in the statement above, condition (2) implies condition (4). In future statements of this sort we omit conditions such as (8) stipulating that collections have the same cardinality.

It will follow from the proof below that if in addition to the hypotheses above, we assume that $\lambda(f_i, f_j) = \mu(f_i) = 0$ for all i, j, then f_i' may be assumed to be regularly homotopic to f_i for each i.

Proof. By hypothesis, $\lambda(g_i, g_j) = 0 = \mu(g_i)$ for all i, j. As a result, all the intersections and self-intersections within G are paired by framed, immersed Whitney discs with pairwise disjoint, embedded boundaries (Proposition 11.10). Similarly, since $\lambda(f_i, g_j) = \delta_{ij}$ for all i, j, all but one intersection point between each f_i and g_i, for each i, as well as all intersections between f_i and g_j, for $i \neq j$, are paired by such Whitney discs. Tube each intersection and self-intersection within F into G using the unpaired intersection points, as shown in Figure 16.2 for a single f. We then obtain a collection of discs, which we still call $F = \{f_i\}$, where $\lambda(f_i, f_j) = \mu(f_i) = 0$ and $\lambda(f_i, g_j) = \delta_{ij}$ for each i, j. The first follows since now all the intersections within F are paired by Whitney discs, obtained as parallel copies of the previous Whitney discs mentioned above. The second follows since all new intersections between F and G are parallels of the intersections within G and are thus algebraically cancelling.

Use the geometric Casson lemma (Lemma 15.3) to make the collections $\{f_i\}$ and $\{g_i\}$ geometrically transverse and call the resulting collections $F' := \{f_i'\}$ and $G' := \{g_i'\}$. Each sphere g_i' is framed since it is regularly homotopic to the framed sphere g_i. Note that the meridian of each f_i' is null-homotopic in the exterior of $\bigcup f_i'$, via the geometrically transverse sphere g_i' punctured at the transverse intersection point. In other words, the collection $\{f_i'\}$ is π_1-negligible, which by definition means that the inclusion induced map is an isomorphism $\pi_1(M \smallsetminus F') \xrightarrow{\cong} \pi_1(M)$, as desired. Since $\pi_1(M)$ is good, $\pi_1(M \smallsetminus F')$ is also good.

Use the fact that $\lambda(f_i', f_j') = \mu(f_i') = 0$ to pair up all the intersections and self-intersections among the $\{f_i'\}$ with framed, immersed Whitney discs with pairwise disjoint, embedded boundaries (Proposition 11.10). Consider one such a Whitney disc D pairing up intersections between f_i' and f_j', where possibly $i = j$. The disc D may intersect itself, the collections $\{f_i'\}$ and $\{g_i'\}$, as well as any number of other Whitney discs (see Figure 16.3). For each intersection of D with f_ℓ', for some ℓ, tube D into a parallel push-off of the geometric dual g_ℓ', as shown in Figure 16.3. This introduces potentially many new intersections between D and anything that intersected g_ℓ' (including g_ℓ' itself), as well as new self-intersections of D coming from the self-intersections of g_ℓ'. However, the interior of D no longer intersects any f_i', since g_ℓ' intersects exactly one f_i', namely f_ℓ', at the intersection point we used for tubing. Repeat this for all the Whitney discs among the intersections of the

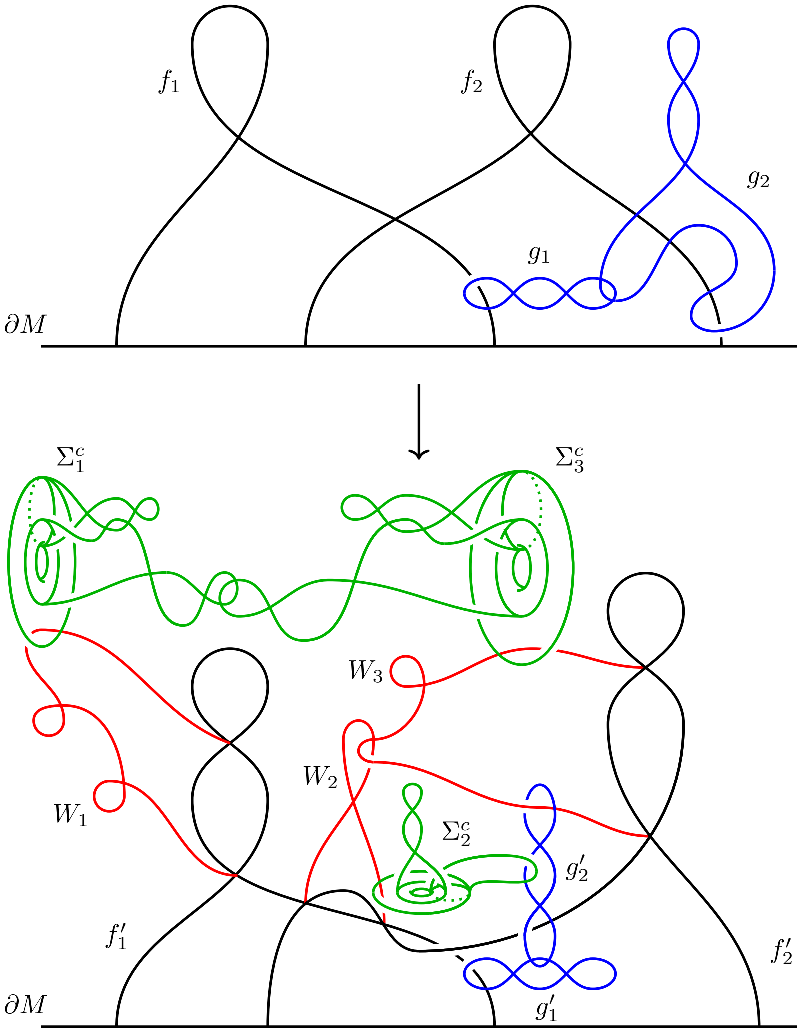


Figure 16.1. Summary of Proposition 16.1. The top depicts the hypotheses of the proposition and the bottom the desired outcome. The body of Σ_k^c lies in a regular neighbourhood of $\{f_i'\}$. Note that the spheres $\{g_i'\}$ may intersect the Whitney discs $\{W_k\}$ or the capped surfaces $\{\Sigma_k^c\}$. The latter intersections occur only with the caps, but we will not need this fact.

$\{f_i'\}$. Now our Whitney discs are more complicated but their interiors lie in the complement of $\bigcup f_i'$ in M. Call this collection of Whitney discs $\{D_k'\}$.

Next we find appropriate geometrically transverse capped surfaces for the $\{D_k'\}$. Note that each D_k' pairs two double points between some f_i' and f_j', where possibly $i = j$. Let Σ_k be the Clifford torus at one of these double points. Note that Σ_k intersects D_k' exactly once, and the collection of such Clifford tori are mutually disjoint, framed, and embedded within a regular neighbourhood of $\bigcup f_i'$. Cap each Σ_k with meridional discs for f_i' and f_j', namely the discs described in Figure 16.4. By construction, the result is a framed, capped surface. Note that each cap has a unique intersection with $\{f_i'\}$, and neither cap intersects $\{D_k'\}$. Tube these intersections into parallel copies of the (framed) geometrically transverse spheres g_i'

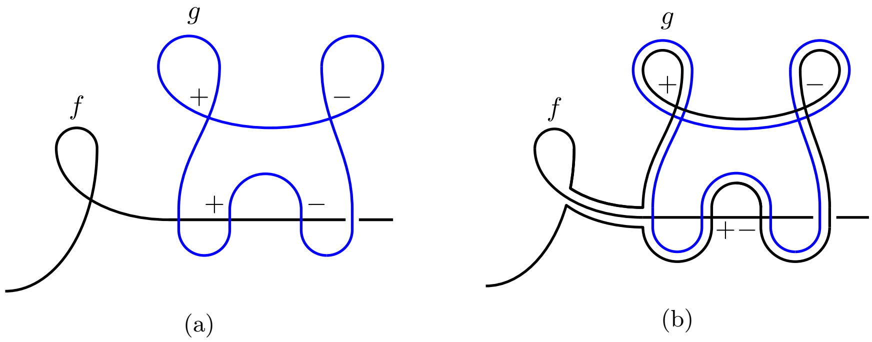


Figure 16.2. Tube each self-intersection of f into g using the unpaired intersection point. After this process, all the self-intersections of f will be paired by Whitney discs, arising as parallel copies of the Whitney discs for the self-intersections of g.

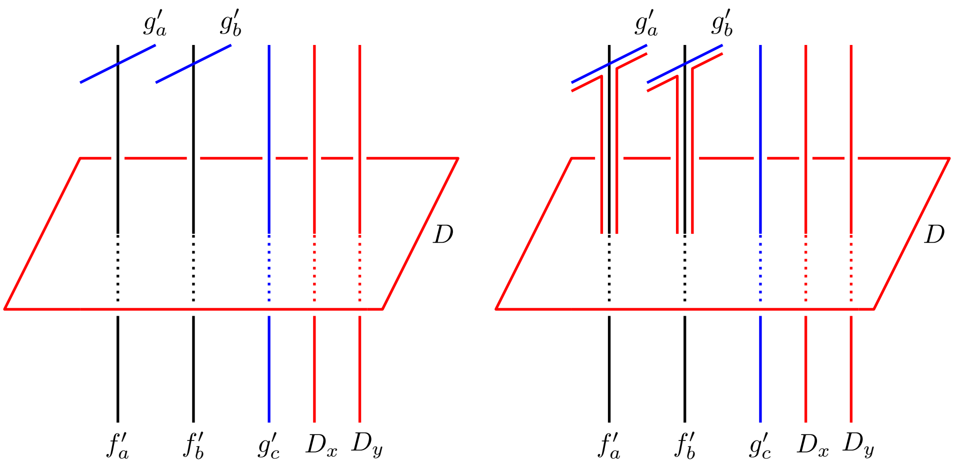


Figure 16.3. Left: A schematic picture of a piece of a Whitney disc D. It may intersect $\{f'_i\}$, $\{g'_i\}$, or other Whitney discs. Note that $\{f'_i\}$ and $\{g'_i\}$ are geometrically transverse. Right: Tube D into the collection $\{g'_i\}$ so that the interior of D lies in $M \smallsetminus \bigcup f'_i$.

and g'_j. Note that if C_ℓ and C_m are any two of these new caps of $\{\Sigma_k\}$, then $\lambda(C_\ell, C_m) = \mu(C_\ell) = 0$ since $\lambda(g'_i, g'_j) = \mu(g'_i) = 0$ for all i, j. Since f'_i and g'_j are disjoint for all $i \neq j$ and $f'_i \pitchfork g'_i$ is the single point used in tubing, the collection of caps is disjoint from $\{f'_i\}$.

The proof of Proposition 16.1 is almost complete. The only problem is that the current caps for $\{\Sigma_k\}$ may intersect the collection of Whitney discs $\{D'_k\}$, due to any intersections between $\{D'_k\}$ and $\{g'_i\}$. Contract a parallel copy of each Σ_k along the current caps, call the result S_k, and push $\{D'_k\}$ off the contraction (do not push off anything else). This produces a new set of Whitney discs $\{W_k\}$, which are equipped with a collection of geometrically transverse spheres $\{S_k\}$ coming from the contraction, and the $\{W_k\}$ pair the intersections and self-intersections within $\{f'_i\}$. Note that these Whitney discs are still contained in $M \smallsetminus \bigcup f'_i$ as needed. Moreover, the collection $\{\Sigma_k\}$, not including the caps, is geometrically transverse

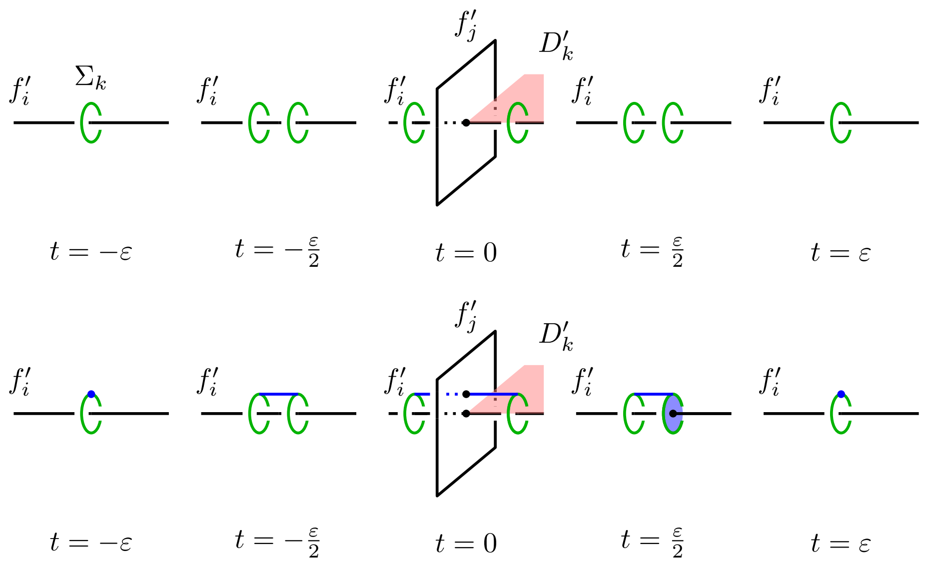


Figure 16.4. Obtaining a geometrically transverse capped surface from a Clifford torus. Top: The Clifford torus Σ_k (green) at one of the two intersection points paired up by the Whitney disc D_k' (red) (compare Figure 15.1). The single point of intersection between Σ_k and D_k' is shown in the middle panel. Bottom: Two meridional discs (blue) are shown and we see that each meridional disc intersects precisely one of f_i' or f_j' precisely once and does not intersect D_k'.

to $\{W_k\}$. Observe also that $\lambda(S_\ell, S_m) = \mu(S_\ell) = 0$ for all ℓ, m since these spheres were produced by contracting capped surfaces with pairwise disjoint and embedded bodies. Tube all the intersections of the caps of $\{\Sigma_k\}$ with $\{W_k\}$ into parallel copies of elements of $\{S_k\}$. The last few steps are shown in Figure 16.5. Cap the collection $\{\Sigma_k\}$ with these new caps, and call the resulting collection $\{\Sigma_k^c\}$. Observe that $\{\Sigma_k^c\}$ is a collection of geometrically transverse capped surfaces for $\{W_k\}$. Since the collection of previous caps, as well as the collection $\{S_k\}$, have trivial intersection and self-intersection numbers, and moreover, the intersections between $\{S_k\}$ and the previous caps are algebraically cancelling, the intersections among the new caps are also algebraically cancelling. That is, $\lambda(C_\ell, C_m) = \mu(C_\ell) = 0$ for any caps C_ℓ and C_m for $\{\Sigma_k^c\}$. This completes the proof. □

Next, we wish to upgrade the Whitney discs obtained in Proposition 16.1 to capped gropes of height two. This is the content of the following proposition. Our choice of notation in the statement of the proposition is intended to help the reader see the connection with the previous proposition, although we will be forced to reuse some symbols within the proofs themselves. The 4-manifold N should be thought of as the manifold $M \smallsetminus \bigcup \nu f_i'$ from before, where $\nu f_i'$ is an small open regular neighbourhood of f_i'.

Proposition 16.2. *Let N be a smooth 4-manifold with nonempty boundary and let*

$$(W_1, \dots, W_q) \colon (D^2 \sqcup \cdots \sqcup D^2, S^1 \sqcup \cdots \sqcup S^1) \looparrowright (N, \partial N)$$

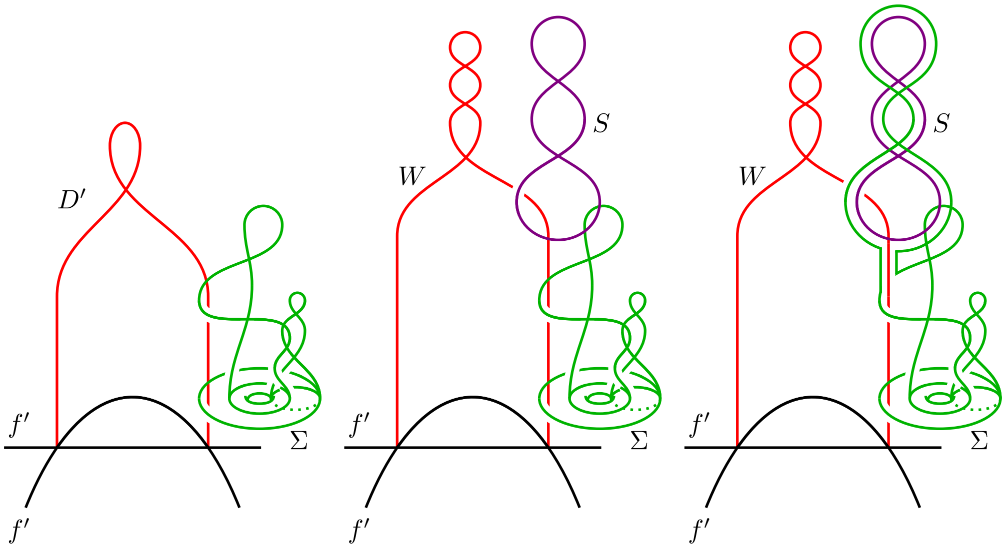


Figure 16.5. Contracting a parallel copy of $\{\Sigma_k\}$ along with the caps produces spheres $\{S_k\}$; pushing $\{D'_k\}$ off the contraction produces new Whitney discs $\{W_k\}$ geometrically transverse to the spheres $\{S_k\}$. Next tube the intersections between the caps of $\{\Sigma_k^c\}$ and $\{W_k\}$ into the spheres $\{S_k\}$. This produces a new set of caps for $\{\Sigma_k\}$.

be a finite immersed collection of discs in N with pairwise disjoint, embedded boundaries. Suppose that $\{\Sigma_k^c\}$ is a collection of geometrically transverse capped surfaces for the $\{W_k\}$ such that

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

for every pair of caps C_ℓ and C_m of $\{\Sigma_k^c\}$. Assume in addition that each Σ_k, namely the body of Σ_k^c, is contained in a collar neighbourhood of ∂N.

Then there exists a collection of disc-like capped gropes $\{G_k^c\}$ of height two embedded in N, whose bodies are pairwise disjoint but whose caps may intersect, such that $\{\Sigma_k^c\}$, after a regular homotopy of the caps, forms a collection of geometrically transverse capped surfaces for the $\{G_k^c\}$. Moreover, for each k, the attaching region of G_k^c coincides with the framed boundary of W_k.

Note that the union $\bigcup G_k^c$ is a union-of-discs-like capped grope of height two. The above proposition is summarised in Figure 16.6, in the case $q = 1$. In the upcoming proof that the geometrically transverse capped surfaces act like 'geometrically transverse sphere factories'. Their purpose is to produce a geometrically transverse sphere whenever we require one. Additionally, it will follow from the proof that the caps of the collection $\{G_k^c\}$ have algebraically cancelling intersections and self-intersections, but this fact will not be useful to us in the sequel.

Proof. We begin with the situation shown in the left panel of Figure 16.6 (shown for a single W). The desired outcome is shown in the right panel of Figure 16.6. Convert each W_k to a surface by tubing any intersections among the $\{W_k\}$, including self-intersections, into parallel copies of the geometrically transverse surfaces $\{\Sigma_k\}$, as shown in Figure 16.7. This produces surfaces $\{W'_k\}$ with the same framed boundary as the $\{W_k\}$. Since $\{\Sigma_k\}$ is a mutually disjoint collection of framed, embedded surfaces, so is $\{W'_k\}$. Use parallel copies of the caps $\{C_\ell\}$ of $\{\Sigma_k\}$ to obtain caps $\{C'_n\}$ for the surfaces $\{W'_k\}$. Note that the interiors of the

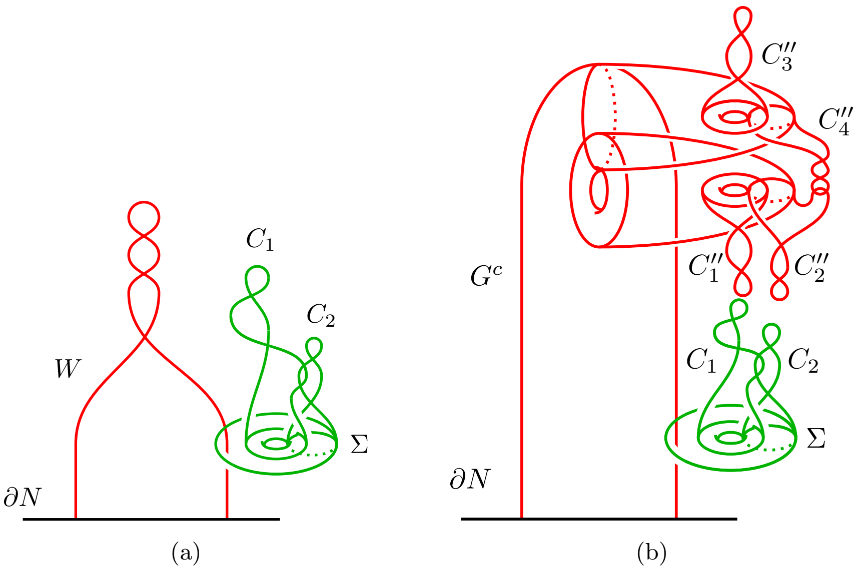


Figure 16.6. Proposition 16.2 goes from the left to the right of the diagram. The caps of the geometrically transverse capped surfaces $\{\Sigma_k^c\}$ may change by a regular homotopy, but the bodies remain the same.

caps $\{C_n'\}$ are disjoint from the surfaces $\{W_k'\}$. Such a capped surface is shown in Figure 16.8.

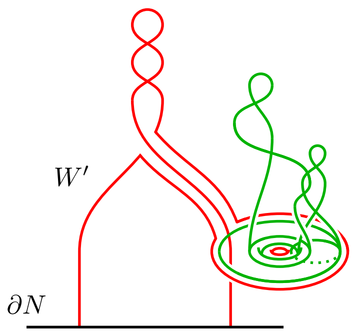


Figure 16.7. Tubing each intersection point of an immersed disc into parallel copies of a geometrically transverse capped surface makes it a grope of height one. Here we show a schematic of the result of tubing a single intersection point.

Now we will separate the caps $\{C_\ell\}$ of the $\{\Sigma_k^c\}$ from the caps $\{C_n'\}$ for the surfaces $\{W_k'\}$. We will need a geometrically transverse sphere for each W_k', so we turn to our geometrically transverse sphere factory: for each k, take a parallel copy of Σ_k^c and contract to obtain a sphere S_k (do not push anything off the contraction), as shown in Figure 16.9. Since the Σ_k are mutually disjoint and embedded such that Σ_k^c intersects $\{W_k'\}$ in a single point on W_k' (and no other element of $\{W_k'\}$), the $\{S_k\}$ constitute a collection of geometrically transverse spheres for the $\{W_k'\}$. Here we are using the fact that the caps $\{C_\ell\}$ and the surfaces $\{W_k'\}$ (not including

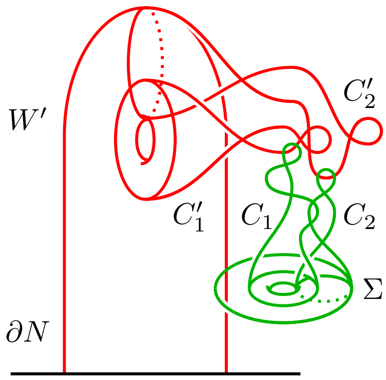


Figure 16.8. A single capped surface W' with a geometrically transverse capped surface Σ^c with body Σ.

the caps) do not intersect. Observe also that $\lambda(S_\ell, S_m) = \mu(S_\ell) = 0$ for all ℓ, m, since these spheres are produced by contracting capped surfaces with mutually disjoint, embedded bodies. Moreover, the $\{S_k\}$ may intersect the caps of the surfaces $\{W_k'\}$ and the caps of the $\{\Sigma_k^c\}$, but these intersections all cancel algebraically.

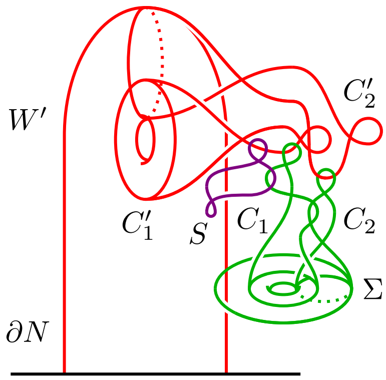


Figure 16.9. The first geometrically transverse sphere S (purple).

Recall that $\lambda(C_\ell, C_m) = \mu(C_\ell) = 0$ for every pair of caps C_ℓ and C_m of the geometrically transverse surfaces $\{\Sigma_k\}$. Since every cap C_n' for some W_k' arises as a parallel copy of a cap for some Σ_k, we see that $\lambda(C', C) = 0$ for any cap C' for some W_k' and any cap C for some Σ_k. Thus the intersection points between C and C' are paired by framed, immersed Whitney discs with pairwise disjoint and embedded boundaries. Let D be such a Whitney disc. We may and shall assume that $D \cap \Sigma_k = \emptyset$ for all k, since D can be pushed off the collar neighbourhood of ∂N containing $\{\Sigma_k\}$. For each intersection of D with some W_k' (not including the caps), tube D into S_k. We obtain a new Whitney disc, which we still call D, having possible intersections only with the caps of $\{\Sigma_k^c\}$ and the caps of $\{W_k'\}$ (the number of such intersections may have increased due to the tubing into S_k), as well as self-intersections. By the proof of the geometric Casson lemma (Lemma 15.3), there is a regular homotopy of the caps making D disjoint from both types of caps at the expense of introducing new cap intersections among the $\{C_\ell\}$ and among the $\{C_n'\}$. That is, push these caps off the correct part of the boundary of D, as shown in Figure 16.10. Although the caps have changed by a regular homotopy, we will continue to refer to them by the previous symbols for convenience.

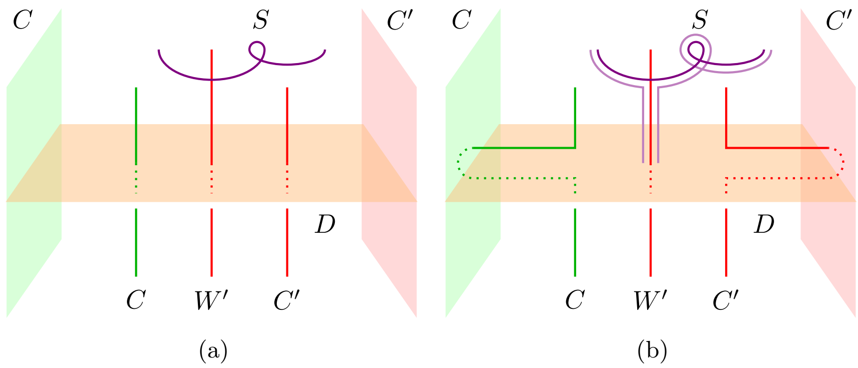


Figure 16.10. Separating the caps of $\{\Sigma_k^c\}$ and $\{W_k'\}$. Tube all the intersections of the Whitney disc D with W' into S, then perform finger moves on C and C' towards C and C' respectively.

A subsequent (immersed) Whitney move of C over D ensures that C' and C are disjoint. There are new (algebraically cancelling) intersections between C and $\{S_k\}$ as a result. Repeat this process to ensure that the caps $\{C_\ell\}$ of the $\{\Sigma_k^c\}$ are pairwise disjoint from the caps $\{C_n'\}$ for the surfaces $\{W_k'\}$, as shown in Figure 16.11.

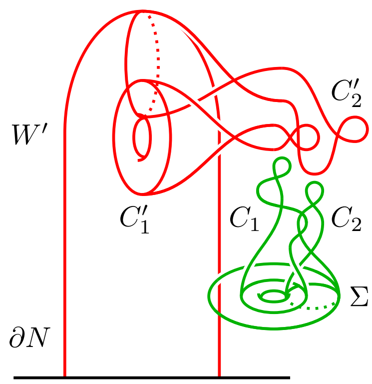


Figure 16.11. After separating the caps of $\{\Sigma_k^c\}$ and $\{W_k'\}$.

Now push down the intersections and self-intersections of the caps $\{C_n'\}$ for the surfaces $\{W_k'\}$ down into the base surface and then tube into parallel copies of the geometrically transverse surfaces $\{\Sigma_k\}$, as shown in Figure 16.12, to promote the collection $\{W_k'\}$ to height two gropes $\{G_k\}$. Observe that the attaching regions of the $\{G_k\}$ coincide with the framed boundary of the $\{W_k\}$. Parallel copies of the caps $\{C_\ell\}$ of the $\{\Sigma_k\}$ provide caps for the $\{G_k\}$, so let $\{G_k^c\}$ denote the capped gropes. However, these caps intersect the caps of the $\{\Sigma_k^c\}$. We will now adjust these to complete the proof.

Remark 16.3. As noted earlier, each S_k may intersect the caps of $\{W_k'\}$ and the caps of $\{\Sigma_k^c\}$. These intersections are algebraically cancelling. By the construction of $\{G_k^c\}$, there are possible (algebraically cancelling) intersections between the bodies of $\{G_k^c\}$ and the spheres $\{S_k\}$. In other words, $\{S_k\}$ is *not* a collection

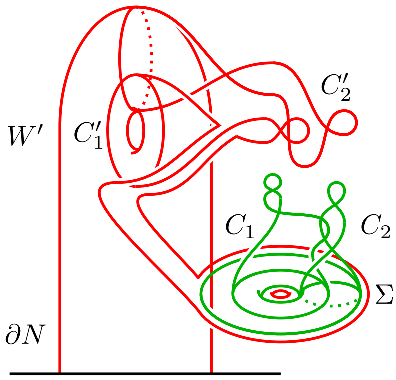


Figure 16.12. Push down the cap intersections of W' into the bottom stage and tube into a push-off of the geometrically transverse capped surface Σ^c with body Σ.

of geometrically transverse spheres for the gropes $\{G_k\}$, even without considering caps. While the collection $\{S_k\}$ is algebraically transverse to the collection of gropes $\{G_k\}$ by construction, our usual technique for upgrading algebraic duals to geometric duals, namely the geometric Casson lemma, could produce undesirable intersections, for instance, in the bodies of the gropes $\{G_k\}$. This is why we use our geometrically transverse sphere factory (namely the geometrically transverse capped surfaces $\{\Sigma_k^c\}$) rather than a single family of geometrically transverse spheres. At multiple points of the proof we produce a geometrically transverse sphere and then use it up, so we need the ability to find a new one each time.

Let $\{C_p''\}$ denote the present collection of caps for the gropes $\{G_k\}$. As noted before, to finish the proof we need only to separate the caps $\{C_\ell\}$ of the transverse capped surfaces $\{\Sigma_k^c\}$ from $\{C_p''\}$, since the collections $\{\Sigma_k\}$ and $\{G_k\}$ are already geometrically transverse. As before, for each k, take a parallel copy of Σ_k along with its caps, and contract to obtain a sphere S_k', as shown in Figure 16.13 (as before, do not push anything off the contraction). Observe as above that each S_k' is a geometrically transverse sphere for G_k and intersects no other $G_{k'}$ for $k \neq k'$. Here we are using the fact that the caps $\{C_\ell\}$ and the gropes $\{G_k\}$ do not intersect. However note that the spheres $\{S_k'\}$ do have (algebraically cancelling) intersections among themselves, coming from the intersections among the caps $\{C_\ell\}$, as well as with the caps of the gropes $\{G_k\}$ and the caps of the $\{\Sigma_k\}$.

Since the caps $\{C_p''\}$ for $\{G_k\}$ were produced as parallel copies of the caps for the geometrically transverse surfaces $\{\Sigma_k\}$, we see that $\lambda(C'', C) = 0$ for every cap C'' for $\{G_k\}$ and every cap C for $\{\Sigma_k\}$. Thus the intersection points between C and C'' are paired by framed, immersed Whitney discs with pairwise disjoint, embedded boundaries. Let D' be such a Whitney disc. As before, $D' \cap \Sigma_k = \varnothing$ for all k. For each intersection of D' with some G_k, tube D' into the geometrically transverse sphere S_k'. We obtain a new Whitney disc, which we still call D', having possible intersections only with the caps of $\{\Sigma_k\}$ and the caps of $\{G_k\}$, possibly more than before due to the tubing into S_k', as well as self-intersections. By the proof of the geometric Casson lemma (Lemma 15.3), there is a regular homotopy making D' disjoint from all the caps, at the expense of creating more cap intersections among the $\{C_\ell\}$ and among the $\{C_p''\}$. To achieve this, push the caps off D' over the correct part of the boundary of D'. A subsequent Whitney move of C over D' reduces the number of intersections between the collections of caps for $\{G_k\}$ and

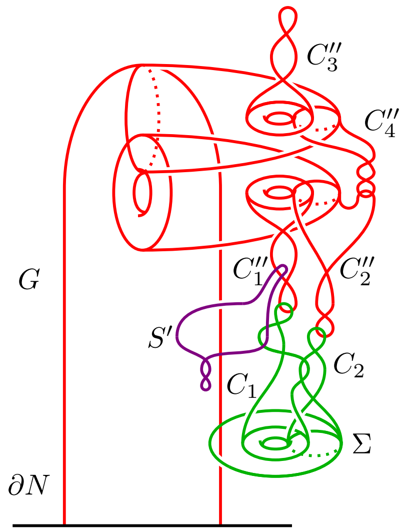


Figure 16.13. The second geometrically transverse sphere S' (purple).

the caps $\{C_\ell\}$ for the geometrically transverse surfaces $\{\Sigma_k\}$. Repeat this process for all the Whitney discs pairing intersection points of $\{C_\ell\}$ and $\{C_p'''\}$, to ensure that the caps for $\{G_k\}$ and the caps $\{C_\ell\}$ for the geometrically transverse surfaces $\{\Sigma_k\}$ are pairwise disjoint, as shown in Figure 16.6(b). Observe that the collection $\{\Sigma_k^c\}$ is geometrically transverse to the capped gropes $\{G_k^c\}$, where in both cases we have used the latest set of caps. In the former collection, the caps differ from the original caps in the hypothesis by a regular homotopy. We have now arranged the situation in Figure 16.6(b) and completed the proof. □

REMARK 16.4. In fact we will only need gropes of height 1.5 for the next step, in Chapter 17. To produce such gropes, we could have merely tubed the intersections of the caps attached to half of a symplectic basis for the homology of the surfaces $\{W_k'\}$ in the proof above. Alternatively, perform a contraction of half of the surfaces in the second storeys of $\{G_k^c\}$. However, this does not decrease the complexity of the proof.

In summary, in this chapter we have replaced the immersed discs $\{f_i\}$ in the hypotheses of the disc embedding theorem by a new family of immersed discs $\{f_i'\}$ with the same framed boundary as the $\{f_i\}$, whose intersections and self-intersections are paired by pairwise disjoint, embedded Whitney circles bounding embedded height two capped gropes $\{G_k^c\}$ (with the correct framed attaching region) equipped with geometrically transverse capped surfaces $\{\Sigma_k^c\}$, all with interiors lying in the complement of $\bigcup f_i'$. We also know that there is a collection $\{g_i'\}$ of framed, geometrically transverse spheres for the $\{f_i'\}$. In the next chapter, we will upgrade the collection $\{G_k^c\}$ to a collection of 1-storey capped towers with geometrically transverse spheres. We will use these new geometrically transverse spheres to make the $\{g_i'\}$ disjoint from the 1-storey capped towers.

CHAPTER 17

Grope height raising and 1-storey capped towers

Peter Feller and Mark Powell

We show how to raise the height of capped gropes. We will apply this procedure to the capped gropes constructed in the previous chapter as well as several other times in the remainder of the proof of the disc embedding theorem.

The goal of this chapter is the following proposition. We give the proof at the end of this chapter, after developing the necessary tools.

PROPOSITION 17.1. *Let N be a smooth, connected 4-manifold with nonempty boundary and with $\pi_1(N)$ good. Let $\{G_k^c\}$ be a collection of disc-like capped gropes of height at least 1.5 in N, whose bodies are mutually disjoint but whose caps might intersect one another, equipped with a collection of geometrically transverse capped surfaces $\{\Sigma_k^c\}$, such that each Σ_k, namely the body of Σ_k^c, is contained in a collar neighbourhood of ∂N.*

Then for every $n \in \mathbb{N}$ there exists a collection of disc-like 1-storey capped towers $\{\mathcal{T}_k^c\}$ embedded in N, whose bodies are mutually disjoint but whose caps may intersect, with first storey gropes of height n and equipped with an immersed collection of framed geometrically transverse spheres $\{R_k\}$. Moreover, $\mathcal{T}_k^c$ and G_k^c have the same attaching region for every k, and each R_k is obtained from Σ_k^c by a contraction.

17.1. Grope height raising

The technique of grope height raising allows us to find an arbitrarily high capped grope inside any capped grope of height at least 1.5 with the same (framed) attaching region. For the convenience of the reader we briefly recall the definition of capped and asymmetric gropes from Section 12.1. In particular, observe that the base stage of a grope may have many components.

DEFINITION 12.6(2). A *capped grope* $G^c(h)$ of height h is a symmetric generalised tower $G_0 \cup G_1 \cup \cdots \cup G_{h+1}$ of height $h+1$ where the stages $G_1, \dots, G_h$ are surface stages and G_{h+1} is a cap stage. The union of the surface stages is called the *body* of the grope.

DEFINITION 12.10. An *asymmetric (capped) grope* is a generalised tower $G = G_0 \cup G_1 \cup G_2 \cup \cdots \cup G_h$, where $G_0 \cup G_1$ is a height one symmetric grope, such that $G_2 \cup \cdots \cup G_h$ has a decomposition as $G^+ \cup G^-$, where G^+ and G^- are symmetric (capped) gropes called the $(+)$-*side* and $(-)$-*side* of G respectively. Let G_0^+ and G_0^- denote the attaching regions of G^+ and G^- respectively. We require that the cores of the attaching regions G_0^+ and G_0^- are isotopic in G_1 to dual halves of a symplectic basis for $H_1(\Sigma_1; \mathbb{Z})$ of curves in the spine Σ_1 of G_1.

If G^+ and G^- are (capped) gropes of heights a and b respectively, we say G has height (a, b).

A grope of height $b + 0.5$ is an asymmetric grope of height $(b, b-1)$.

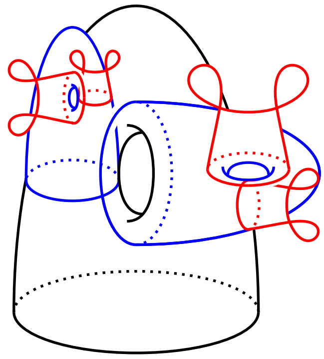


Figure 17.1. Schematic picture of the 2-dimensional spine of a height two capped grope.

In the previous two definitions gropes can be union-of-discs-like or union-of-spheres-like. We also introduce the following terminology.

Definition 17.2. Let $\mathcal{A}$ and $\mathcal{B}$ be generalised towers such that $\mathcal{A} \subseteq \mathcal{B}$. Let $m \geq 1$. We say that *the first m stages of $\mathcal{A}$ and $\mathcal{B}$ coincide*, or are *the same*, if the following holds.

(1) The spines of the first m stages of $\mathcal{A}$ and $\mathcal{B}$ coincide,
(2) the inclusion of the first m stages of $\mathcal{A}$ into $\mathcal{B}$ is isotopic to a diffeomorphism with image the first m stages of $\mathcal{B}$, and
(3) the (framed) attaching regions of $\mathcal{A}$ and $\mathcal{B}$ coincide.

If $\mathcal{A}$ and $\mathcal{B}$ are both capped gropes or towers with height m and the first m stages coincide, we say that $\mathcal{A}$ and $\mathcal{B}$ *have the same body.*

We leave it to the reader to determine the correct definition for when $m \in \frac{1}{2}\mathbb{N}$ and $\mathcal{A}$ and $\mathcal{B}$ are asymmetric generalised towers.

Here is the precise statement of grope height raising. This applies to both union-of-discs-like and union-of-spheres-like gropes.

Proposition 17.3 (Grope height raising). *Given a union-of-discs-like or union-of-spheres-like capped grope $G^c(m)$ of height $m \in \frac{1}{2}\mathbb{Z}$, where m is at least 1.5, and a positive integer $n \geq m$, there exists a capped grope $G^c(n)$ of height n embedded in $G^c(m)$ such that the first m stages of $G^c(n)$ and $G^c(m)$ coincide.*

Moreover, any collection of geometrically transverse capped gropes or spheres for $G^c(m)$ is also geometrically transverse to $G^c(n)$.

The proof of grope height raising will combine the following two results.

Lemma 17.4 (Cap separation lemma). *Given an asymmetric union-of-discs-like or union-of-spheres-like capped grope $G^c(m)$ of height (a, b), where $a \geq 1$, there exists a capped grope embedded in $G^c(m)$, with the same height, body, and attaching region as $G^c(m)$, such that its $(+)$-side caps are disjoint from its $(-)$-side caps.*

This lemma first appeared in print in [**CP16**].

Proof. First we construct a union-of-spheres-like capped grope T^c_- such that its body T_- and the $(-)$-side $G^c(m)^-$ of $G^c(m)$ are geometrically transverse. Note that $G^c(m)^-$ might consist only of caps. Take two parallel push-offs of the $(+)$-side grope $G^c(m)^+$. The induced push-offs of the attaching circles of the $(+)$-side grope $G^c(m)^+$ are connected by annuli which intersect the bottom stage of the $(-)$-side grope $G^c(m)^-$ once transversely, as shown in Figure 17.2. Let T^c_- be the union of the two parallel push-offs and the corresponding annuli. By construction, we see that the body of the union-of-spheres-like grope T^c_- and the $(-)$-side grope $G^c(m)^-$ are geometrically transverse, while the caps of T^c_- may intersect one another or the caps of $G^c(m)$. By the height assumption on $G^c(m)$, T^c_- has height at least one.

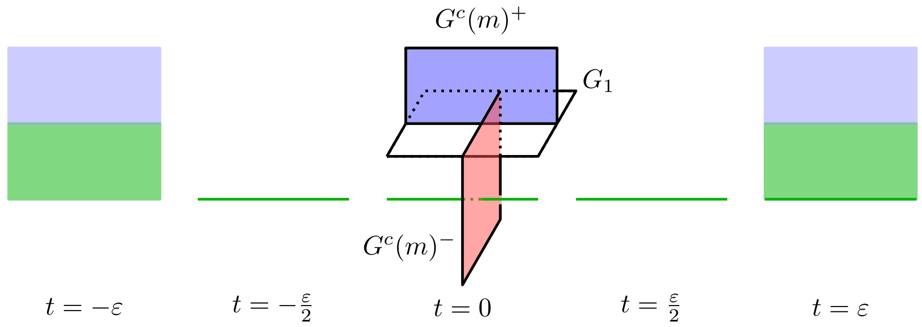


Figure 17.2. We show a local picture for the attaching circle of a single component of the body of the $(+)$-side $G^c(m)^+$ (blue) to a first stage surface G_1 of $G^c(m)$. The $(-)$-side is shown in red. Lighter blue depicts the two parallel push-offs of the $(+)$-side, while green indicates an annulus connecting the two push-offs of the attaching circle. Note the single transverse intersection between the annulus and the $(-)$-side in the central image.

Contract T^c_- to produce an immersed collection of spheres $\{S^i_-\}$ and push all the $(-)$-side caps off the contraction, at the cost of introducing more (algebraically cancelling) intersections among the $(-)$-side caps. At this point, the body and the $(+)$-side caps of $G^c(m)$ remain the same, the $(-)$-side caps may have changed, and the collection of components of the bottom stage of the $(-)$-side capped grope has a collection of geometrically transverse spheres $\{S^i_-\}$. These spheres may intersect the $(+)$-side caps but are disjoint from the $(-)$-side caps. In fact the intersections between $\{S^i_-\}$ and the $(+)$-side caps are algebraically cancelling since the $\{S^i_-\}$ arose from contractions, but we will not need this fact.

Finally, push all the intersections of the $(+)$-side caps with the $(-)$-side caps down to the bottom stage of the $(-)$-side and tube into parallel push-offs of $\{S^i_-\}$ (see Figure 17.3). Of course, if the bottom stage of the $(-)$-side consists only of caps, such as when $G^c(m)$ has height 1.5, then no pushing down is needed. This achieves the separation of $(-)$-side caps and $(+)$-side caps; the grope $G^c(m)$ has new $(+)$-side caps and its $(-)$-side caps have been changed by a regular homotopy, but remains unchanged otherwise. □

Lemma 17.5. *Given a union-of-discs-like or union-of-spheres-like capped grope $G^c(a,b)$ of height (a,b), where $a \geq 1$, there exists a capped grope $G^c(a,b+a)$ of height $(a,b+a)$ embedded in $G^c(a,b)$, with the same first stage and attaching region, such that the bodies of the $(+)$-side gropes and the first b stages of the $(-)$-side gropes of $G^c(a,b)$ and $G^c(a,b+a)$ coincide.*

Proof. Apply the cap separation lemma (Lemma 17.4) to arrange that the $(+)$-side caps of $G^c(a,b)$ are disjoint from the $(-)$-side caps. Let T^c_- be a union-of-spheres-like capped grope of height a transverse to the bottom stage of the $(-)$-side, constructed as in the first step of the proof of Lemma 17.4. As before, the body of T^c_- is disjoint from $G^c(a,b)$ aside from the transverse intersections with the bottom stage of the $(-)$-side while the caps of T^c_- only intersect the $(+)$-side caps of $G^c(a,b)$;

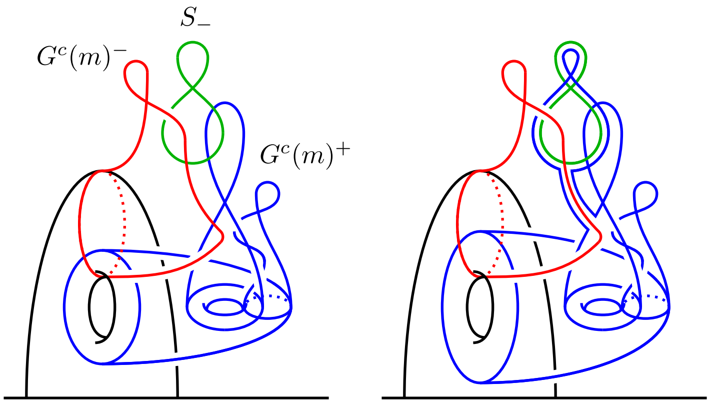


Figure 17.3. Separating $(+)$-side and $(-)$-side caps. The $(+)$-side is shown in blue and the $(-)$-side in red. The immersed sphere S_- was produced in the previous step of the proof.

the latter follows from the fact that the set of $(+)$-side caps and the set of $(-)$-side caps of $G^c(a, b)$ are disjoint from each other.

Push the intersections among the $(-)$-side caps down to the bottom stage of the $(-)$-side and tube into parallel push-offs of T^c_-; this turns each $(-)$-side cap into a capped grope of height a whose attaching region coincides with that of the $(-)$-side caps (see Figure 17.4). Replace the $(-)$-side caps of $G^c(a, b)$ with these new gropes, which become part of the $(-)$-side, to obtain the grope $G^c(a, b + a)$. □

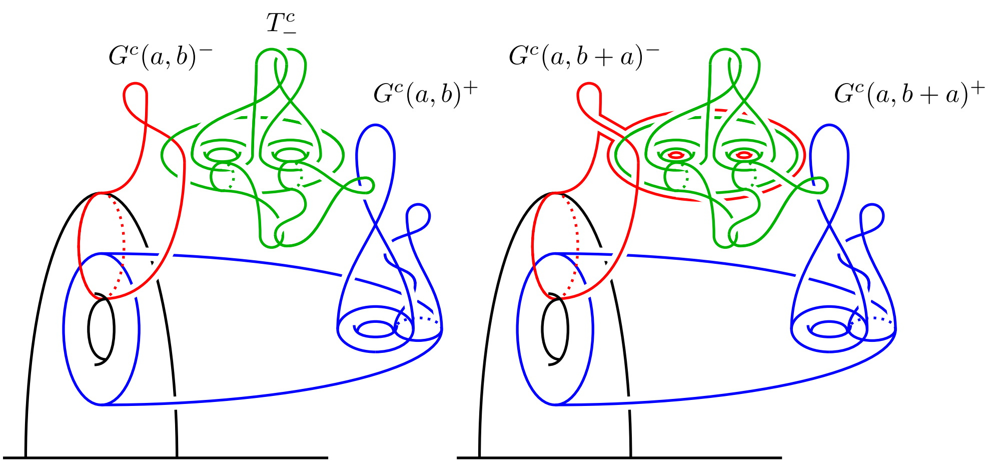


Figure 17.4. Raising the height of the $(-)$-side of $G^c(a, b)$. As before the $(+)$-side is shown in blue and the $(-)$-side in red. Tube the intersections of the $(-)$-side caps into the transverse grope T^c_-. Use parallel copies of the caps for T^c_- to obtain caps for the new $(-)$-side.

Next we give the proof of grope height raising.

Proof of Proposition 17.3 (Grope height raising). We start with an asymmetric capped grope of height $m \geq 1.5$. In other words, this is a capped grope of height (a_0, b_0), where $(a_0, b_0) = (m - 1, m - 1)$ if m is an integer and

$(a_0, b_0) = (\lfloor m \rfloor, \lfloor m \rfloor - 1)$ if m is only a half integer. Since $m \geq 1$, we have $a_0 \geq 1$ in both cases.

A first application of Lemma 17.5 to $G^c(a_0, b_0) := G^c(m)$ yields a capped grope $G^c(a_0, b_0 + a_0)$ of height $(a_0, b_0 + a_0)$. Switch the labels of the $(+)$- and $(-)$-sides, apply Lemma 17.5 again, and switch the labels back. This produces a capped grope $G^c(a_0 + (b_0 + a_0), b_0 + a_0)$ of height $(a_0 + (b_0 + a_0), b_0 + a_0)$. Continue iteratively to find a sequence of capped gropes $\{G^c(a_i, b_i)\}$ of height (a_i, b_i) where the integers a_i and b_i are recursively given as follows for $i \geq 0$.

$$\begin{aligned} &a_{2i+1} = a_{2i}, && b_{2i+1} = a_{2i} + b_{2i}, \\ &a_{2i+2} = a_{2i+1} + b_{2i+1}, && b_{2i+2} = b_{2i+1}. \end{aligned} \tag{17.1}$$

Take a capped grope $G^c(a', b')$ in this sequence of height (a', b'), with $a', b' \geq n-1$, and contract both the $(+)$-side and the $(-)$-side until they each have height $n-1$. This produces a capped grope $G^c(n)$ of height n (that is, height $(n-1, n-1)$) as desired.

For the final sentence in the statement, note that our construction took place in a regular neighbourhood of the second and higher stages of the original grope $G^c(m)$. Since any transverse collection of capped gropes or spheres only intersects the bottom stage of $G^c(m)$ and is disjoint from all other stages (including caps), the statement follows. □

17.2. 1-storey capped towers

In the previous section we saw how to increase the height of certain gropes. We now wish to upgrade gropes to 1-storey capped towers. Recall that, roughly speaking, a 1-storey capped tower consists of a capped grope equipped with mutually disjoint caps with algebraically cancelling double points and a second layer of caps bounded by the double point loops of the caps of the capped grope. This is the only step in the proof of the disc embedding theorem where we will need the good group hypothesis. We recall the definition of good groups for the convenience of the reader.

Definition 12.12. A group Γ is said to be *good* if for every disc-like (i.e. connected body with a single boundary component) capped grope $G^c(1.5)$ of height 1.5 with some choice of basepoint, and every group homomorphism $\phi\colon \pi_1(G^c(1.5)) \to \Gamma$, there exists an immersed disc $D^2 \looparrowright G^c(1.5)$, whose framed boundary coincides with the attaching region of $G^c(1.5)$, such that the elements in $\pi_1(G^c(1.5))$ given by its double point loops, considered as fundamental group elements by making some choice of basing path, are mapped to the identity element of Γ by ϕ.

We will use the following notion in the next proposition as well as in the next chapter.

Definition 17.6. A subset G^c (usually a capped grope) of a 4-manifold M is said to be π_1-*null* in M if all maps of S^1 to G^c are null-homotopic as maps to M. In other words, the inclusion induced map $\pi_1(G^c) \to \pi_1(M)$ is trivial.

The following lemma will be helpful again in the next chapter. This lemma applies to both union-of-discs-like and union-of-spheres-like gropes.

Lemma 17.7 (Sequential contraction lemma). *For $m \geq 0$, let $G^c(m+1)$ be a height $(m+1)$ union-of-discs-like or union-of-spheres-like capped grope. There exists an embedded height m capped grope $G^c(m)$ in $G^c(m+1)$, where the caps of $G^c(m)$ are mutually disjoint and have algebraically cancelling double points, such that the body of $G^c(m)$ coincides with the first m stages of $G^c(m+1)$.*

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

The following proposition upgrades certain gropes with geometrically transverse capped surfaces to 1-storey capped towers with geometrically transverse spheres and the same attaching region. Note that this is the first time in the proof of the disc embedding theorem that the properties of the ambient 4-manifold have become relevant. We found Whitney discs using the hypotheses on the intersection numbers λ and μ (Proposition 11.10) and since then all constructions have taken place within a neighbourhood of the given or constructed surfaces. This will also continue afterwards, but to find the second layer of caps necessary to construct a capped tower, we need to use the fundamental group data of the ambient 4-manifold.

Proposition 17.8. *Let $n \in \mathbb{N}$ and let N be a smooth, connected 4-manifold with nonempty boundary. Let $G^c(m)$ be a union-of-discs-like or union-of-spheres-like capped grope of height m in N, with a collection of geometrically transverse capped surfaces $\{\Sigma_k^c\}$, such that each Σ_k, namely the body of Σ_k^c, is contained in a collar neighbourhood of ∂N. Assume either that*

(1) $\pi_1(N)$ is good and $m = n + 3.5$, or
(2) $G^c(m)$ is π_1-null in N and $m = n + 1$.

Then, corresponding to the type of $G^c(m)$, there exists a union-of-discs-like or union-of-spheres-like 1-storey capped tower $\mathcal{T}^c$ embedded in N whose first storey grope has height n, such that the first n stages of $\mathcal{T}^c$ and $G^c(m)$ coincide, equipped with an immersed collection of geometrically transverse, framed spheres $\{R_k\}$. Moreover, each R_k is obtained from Σ_k^c by a contraction.

Proof. First we address case (1). Apply the sequential contraction lemma (Lemma 17.7) to produce a capped grope $G^c(n+2.5)$ of height $(n+2.5)$ with mutually disjoint caps. Consider the union of the top 1.5 stages of $G^c(n+2.5)$. Since $\pi_1(N)$ is good, each component (namely a disc-like capped grope of height 1.5) contains an immersed disc, whose framed boundary coincides with the attaching region, such that the double point loops are null-homotopic in N.

Remark 17.9. We observe that for the previous step, in fact all we needed is that the image of the inclusion induced map $\pi_1(G^c(m)) \to \pi_1(N)$ is good.

Attach these immersed discs to the lower stages, producing a capped grope $G^c(n+1)$ of height $n+1$ with mutually disjoint caps. Contract the top stage. This results in a new capped grope $G^c(n)$ of height n whose caps are mutually disjoint and have algebraically cancelling double points. Moreover, the double point loops are parallel push-offs of the double point loops for the cap intersections of $G^c(n+1)$ and are thus also null-homotopic. Null homotopies produce immersed discs $\{\widetilde{\delta}_\alpha\}$ bounded by the double point loops of the caps of $G^c(n)$. These discs do not intersect $\{\Sigma_k\}$, namely the collection of bodies of $\{\Sigma_k^c\}$, since the latter set is located in a collar neighbourhood of ∂N and any null homotopy may be pushed off ∂N. On the other hand the discs coming from null homotopies might intersect $G^c(n)$ or the caps of $\{\Sigma_k^c\}$, as shown in Figure 17.5.

We take a short interlude to address case (2). Apply the sequential contraction lemma (Lemma 17.7) to produce a capped grope $G^c(n)$ of height n with mutually disjoint caps and algebraically cancelling double points. Since $G^c(n)$ lies within $G^c(m)$, which is π_1-null in N, the grope $G^c(n)$ is also π_1-null in N. Thus, the double point loops of $G^c(n)$ are null-homotopic in N, and the null homotopies provide immersed discs $\{\widetilde{\delta}_\alpha\}$ bounded by the double point loops of the caps of $G^c(n)$. As

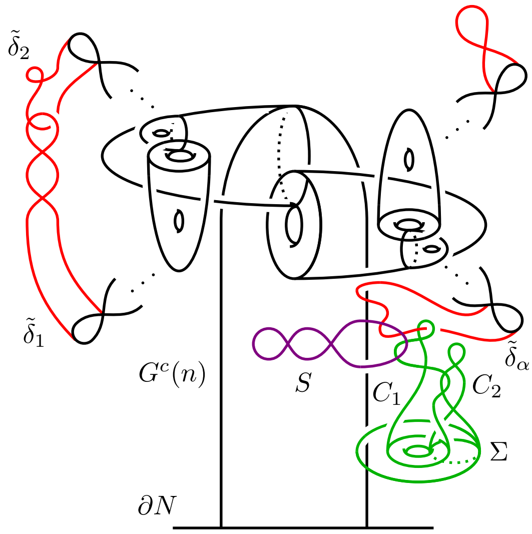


Figure 17.5. Obtaining discs from null homotopies of the double point loops. The geometrically transverse sphere S is obtained by contracting a parallel copy of the geometrically transverse capped surface.

in the previous paragraph, these discs do not intersect $\{\Sigma_k\}$, since the latter set is located in a collar neighbourhood of ∂N, but they might intersect $G^c(n)$ or the caps of $\{\Sigma_k^c\}$, as shown in Figure 17.5.

The remainder of the proof applies to case (1) and case (2). Take a parallel copy of each Σ_k along with its caps and contract, to produce a family of spheres $\{S_k\}$, as shown in Figure 17.5 (do not push anything off the contraction). Boundary twist the discs $\{\widetilde{\delta}_\alpha\}$ to achieve the correct framing (see Remark 17.10 below) and then push down and tube into $\{S_k\}$ to remove any intersections of the resulting discs with $G^c(n)$, as shown in Figure 17.6. Thicken the resulting discs $\{\delta_\alpha\}$ to produce cap stages, and then glue the resulting caps to $G^c(n)$ to produce the 1-storey capped tower $\mathcal{T}^c$.

Remark 17.10. We pause the proof for a moment to address the framing for the tower caps just obtained. First, the subtlety we are about to describe may be circumvented by adding trivial local cusps to the interior of the discs $\{\widetilde{\delta}_\alpha\}$ so that each has algebraically zero double points. Then we need only boundary twist until the framing induced by the normal bundle of $\{\widetilde{\delta}_\alpha\}$, and an appropriate choice of orientation on the fibres, coincides with the framing of the tip region of $G^n(c)$, and continue with the proof.

For the reader who wishes for subtlety, we show how to obtain cap blocks without assuming algebraically zero double points. First we recall some details about cap blocks. By definition (see Definition 12.3) the attaching region of a cap block is framed as the attaching region of the original $D^2 \times D^2$ prior to any plumbing being performed. In Section 13.4.2, we saw that this translates into the Seifert framing for the attaching circle of a cap block in the corresponding Kirby diagram. However, this framing on the attaching region of the cap block coincides with the framing induced by the normal bundle of the spine if and only if the number of self-intersections of the core disc is algebraically zero. More precisely, the two framings differ by twice the number of self-intersections of the core disc. Recall

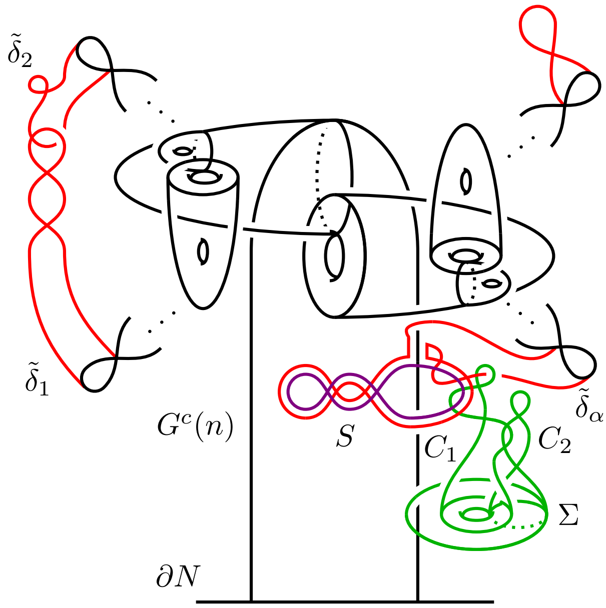


Figure 17.6. After boundary twisting to correct the framing, tube into the geometrically transverse sphere S to remove the intersections of $\{\widetilde{\delta}_\alpha\}$ with the body of $G^c(n)$. Note that $\{\widetilde{\delta}_\alpha\}$ can intersect $G^c(n)$ at any surface stage.

that a cap block is a self-plumbed thickened disc, and includes the information of both an orientation on the core immersed disc and an orientation on the thickening. Consequently, in contrast with the discussion at the beginning of Chapter 11, the orientation data associated with a cap block determines an orientation on the fibres of the thickening, so we do not need to specify this as auxiliary data in order to obtain an induced framing.

For the current construction, our goal is to improve the immersed discs $\{\widetilde{\delta}_\alpha\}$ into cap blocks, endowed in particular with orientation and attaching region data. The boundary of an immersed disc $\widetilde{\delta}_\alpha$ and the core of a component of a tip region of $G^c(n)$ coincide. The tip region is framed. We wish to perform boundary twisting on $\widetilde{\delta}_\alpha$ so that a thickening of the resulting disc is a cap block attached to this tip region with the right framing.

First we discuss orientations. Orient the fibres of the normal bundle of $\widetilde{\delta}_\alpha$ so that on $\partial\widetilde{\delta}_\alpha$ the orientations agree with the orientations of the fibres coming from the parametrisation of the tip region of $G^c(n)$. Together with a choice of orientation of $\widetilde{\delta}_\alpha$, this determines an orientation of the thickening, so after making this choice we will have the orientation data required for a cap block. Make the choice of orientation of each $\widetilde{\delta}_\alpha$ so that the union of $G^c(n)$ and the thickening of the $\{\widetilde{\delta}_\alpha\}$ is an oriented 4-manifold.

Now we show that the correct framings can be arranged, which by the choices made in the previous paragraph will all correspond to the same orientation on the fibres. The framing Y on the boundary that must be used, in order for the thickened boundary to be the attaching region of a cap block, differs from the framing Z induced by the normal bundle of $\widetilde{\delta}_\alpha$ by a number of full twists equal to twice the count of self-intersections e. Since framings are affine over $\mathbb{Z}$, heuristically we write $Y - Z = 2e$. Both Y and Z may change under boundary twisting, to Y' and Z'

say, but since boundary twisting does not change the number of self-intersections of $\widetilde{\delta}_\alpha$ they change by the same amount, so $Y - Z = Y' - Z' = 2e$. Moreover for a single boundary twist we have $Z' - Z = \pm 1$, so boundary twisting has a nontrivial effect on both Y and Z.

Now to obtain a cap block and thence a 1-storey capped tower, we perform boundary twisting on $\widetilde{\delta}_\alpha$, changing the framings Y and Z, respectively to Y' and Z' as just described, so that the framing X on the tip region of $G^c(n)$ differs from the framing Z' induced by the normal bundle of the final disc by twice the count of self-intersections of $\widetilde{\delta}_\alpha$. In symbols this is $X - Z' = 2e$. We may arrange this since X is unchanged by a boundary twist, and every integer can be realised as $Z' - Z$. The final tubing into the framed spheres $\{S_k\}$ does not change the framing. Since also $Y' - Z' = 2e$ we deduce that $X = Y'$. In other words the framings of the tip and attaching regions coincide, and so we have the cap block we desire.

Returning to the proof of Proposition 17.8, note that at this point the caps for $\{\Sigma_k^c\}$ only (possibly) intersect $\mathcal{T}^c$ in the tower caps. Contract each Σ_k^c along its caps and call this family of spheres $\{R_k\}$. The $\{R_k\}$ are framed spheres since they arose from contraction of a capped surface with framed caps. Push all intersections with tower caps off the contraction. This produces more intersections among the tower caps of $\mathcal{T}^c$. The family of spheres $\{R_k\}$ are then geometrically transverse to the resulting 1-storey capped tower as desired. We see the end result of this procedure in Figure 17.7 for a capped tower with connected base surface stage. Observe that the tower caps can intersect one another without restriction, and thus in general we only obtain a union-of-discs-like or union-of-spheres-like 1-storey capped tower rather than a mutually disjoint union of disc-like or sphere-like 1-storey capped towers. □

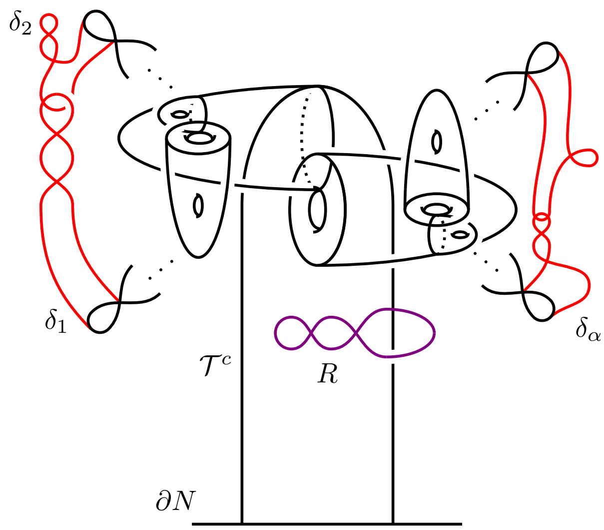


Figure 17.7. A 1-storey capped tower with a geometrically transverse sphere. Note that the immersed discs (black) forming the terminal disc stage of the tower, are not shown completely (each has algebraically cancelling double points).

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

Lemma 17.11. *Let M be a smooth 4-manifold with nonempty boundary and 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 boundaries, that admits an immersed collection $G' := \bigsqcup_{i=1}^n g_i'$ of framed geometrically dual spheres in M. Let $\Sigma^c \subseteq N := M \smallsetminus F'$ be a capped surface constructed by taking a Clifford torus corresponding to an intersection point between f_i' and f_j', where $i = j$ is permitted, equipped with caps obtained by tubing meridional discs for f_i' and f_j' into parallel copies of g_i' and g_j'. Let $R \colon S^2 \looparrowright N$ be the 2-sphere obtained by contracting Σ^c along its caps. Then $\iota_*[R] = 0 \in \pi_2(M)$, where $\iota \colon N = M \smallsetminus F' \to M$ is the inclusion map.

Proof. By Lemma 15.2, and as explained in [**FQ90**, Section 2.3], the homotopy class of a 2-sphere obtained by contracting a torus along caps is independent of the choice of caps, provided the boundaries of the different choices of caps coincide. Therefore in N we may replace the caps constructed from g_i' and g_j' by the meridional caps. The sphere R' resulting from contraction along the meridional caps is contained in a D^4 neighbourhood in M of the intersection point giving rise to the Clifford torus. So R' is null-homotopic in M. It follows that $\iota \circ R$ is also null-homotopic as desired. $\square$

Lemma 17.11 implies in particular that each of the spheres $\{R_k\}$ in Proposition 17.8 is null-homotopic in M, if the $\{\Sigma_k\}$ arise as Clifford tori, as we will see in the next section.

17.3. Continuation of the proof of the disc embedding theorem

Let us return to the proof of the disc embedding theorem. First we prove the proposition stated at the beginning of this chapter.

Proof of Proposition 17.1. As noted earlier $G^c(m) := \bigcup G_k^c$ is a union-of-discs-like capped grope with height $m \geq 1.5$. Apply grope height raising (Proposition 17.3) or contraction to obtain a grope of height $n + 3.5$ and then apply Proposition 17.8. Note that only the last step required the hypothesis that $\pi_1(N)$ be good. $\square$

The combination of our work so far in Part II yields the following proposition.

Proposition 17.12. *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 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 an immersed collection

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

of discs in M such that the following holds.

1. *For each i, the discs f_i and f_i' have the same framed boundary.*
2. *There is an immersed collection*

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

of framed geometrically transverse spheres for F' in M, such that for every i, the sphere $\overline{g}_i$ is homotopic to g_i.

(3) *The intersections and self-intersections within F' are algebraically cancelling.*

(4) *Whitney circles for the intersections and self-intersections within F' bound a union-of-discs-like* 1*-storey capped tower, with arbitrarily many surface stages, that apart from its attaching region lies in the complement in M of both F' and $\overline{G}$, and such that the framing of the attaching region agrees with the Whitney framing.*

Proof. The notation in our proposition statements has been uniform, to aid with splicing them together now. Proposition 16.1 begins with the same hypotheses as the current statement, and produces an immersed collection $F' = \{f'_i\}$ of discs, with the same framed boundary as F and with geometrically dual spheres $G' = \{g'_i\}$. Additionally, the intersections and self-intersections of F' are paired by Whitney discs $\{W_k\}$ equipped with geometrically transverse capped surfaces $\{\Sigma_k^c\}$. The bodies $\{\Sigma_k\}$ of the $\{\Sigma_k^c\}$ lie in a regular neighbourhood of F' and the caps have algebraically cancelling intersections and self-intersections. Finally the spheres g'_i and g_i are regularly homotopic for each i.

Applying Proposition 16.2 to these data, using $N := M \smallsetminus \bigcup \nu f'_i$, upgrades the discs $\{W_k\}$ to height two disc-like capped gropes $\{G_k^c\}$, whose attaching region coincides with the framed boundary of $\{W_k\}$, and such that $\{\Sigma_k^c\}$, after a regular homotopy of the caps, is geometrically transverse to $\{G_k^c\}$. Due to the existence of the geometrically dual spheres for F', the group $\pi_1(N)$ is good. The bodies of $\{G_k^c\}$ are mutually disjoint but the caps may intersect.

Now apply Proposition 17.1 to $\{G_k^c\}$ and $\{\Sigma_k^c\}$ in N to produce a collection of disc-like 1-storey capped towers $\{\mathcal{T}_k^c\}$ with the same attaching region as $\{G_k^c\}$. As before, the bodies of $\{\mathcal{T}_k^c\}$ are mutually disjoint but the tower caps may intersect. That is, the $\{\mathcal{T}_k^c\}$ form a union-of-discs-like 1-storey capped tower. We also obtain the family of spheres $\{R_k\}$ geometrically dual to $\{\mathcal{T}_k^c\}$. Applying Lemma 17.11 using the discs F', the spheres G', and the capped surfaces $\{\Sigma_k^c\}$ in M, we see that each R_k is null-homotopic in M.

The spheres $G' = \{g'_i\}$ produced using Proposition 16.1 may intersect the 1-storey capped towers $\{\mathcal{T}_k^c\}$. If some g'_i intersects some $\mathcal{T}_k^c$, push down the intersection and tube into R_k, as shown in Figure 17.8. Note that an intersection between g'_i and $\mathcal{T}_k^c$ may occur at any surface, disc, or cap stage of $\mathcal{T}_k^c$, so we may need to push down the intersection several times before reaching the base surface. Call the resulting spheres $\overline{G} := \{\overline{g}_i\}$. Since the $\{R_k\}$ are framed, so are the $\{\overline{g}_i\}$. Now the collection $\{\mathcal{T}_i^c\}$ lies in the complement of $\{f'_i\}$ and $\{\overline{g}_i\}$, apart from its attaching region, as desired. Moreover, the families $\{f'_i\}$ and $\{\overline{g}_i\}$ are geometrically transverse. Since each R_k is null-homotopic in M, the sphere $\overline{g}_i$ is homotopic to g'_i in M for each i. We saw earlier that each g'_i is homotopic to g_i. This completes the proof. □

In summary, Proposition 17.12 starts with the hypotheses of the disc embedding theorem and produces immersed discs $\{f'_i\}$, along with a collection of framed, geometrically transverse spheres $\{\overline{g}_i\}$, whose intersections and self-intersections are paired by a union-of-discs-like 1-storey capped tower lying in the complement of both $\{f'_i\}$ and $\{\overline{g}_i\}$.

The rest of the proof of the disc embedding theorem has two steps. First we will show, in Chapter 18, that any union-of-discs-like 1-storey capped tower with at least four surface stages contains a pairwise disjoint union of embedded skyscrapers with the same attaching region (Theorem 18.9). Then, in Part IV, we will show that any skyscraper is homeomorphic to the standard 2-handle, relative to the attaching region (Theorem 27.1). Thus, in Proposition 17.12, we can replace the 1-storey capped tower by a pairwise disjoint collection of flat, embedded, and framed Whitney discs for the intersections and self-intersections of $\{f'_i\}$. Performing the

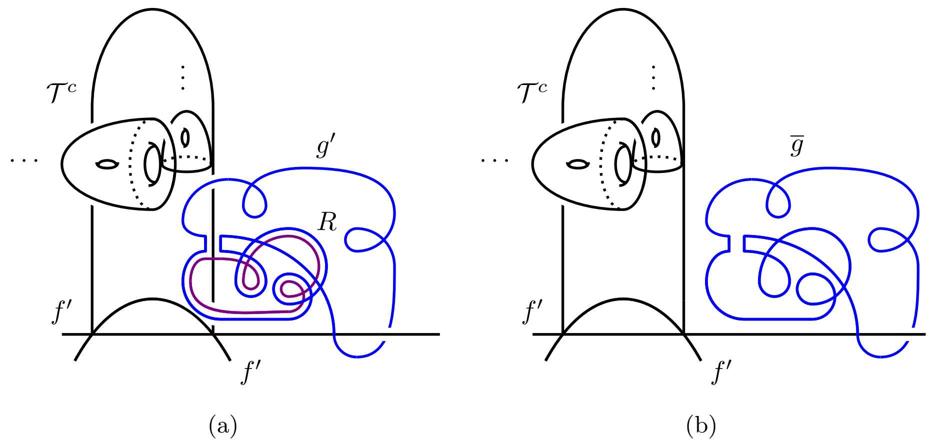


Figure 17.8. Tube to remove intersections of $\{g_i'\}$ with $\{\mathcal{T}_k^c\}$, ensuring that the collection $\{\mathcal{T}_k^c\}$ lies in the complement of $\{f_i'\}$ and $\{\overline{g}_i\}$, other than the attaching regions.

Whitney move over these Whitney discs produces the embedded discs we desire. The spheres $\{\overline{g}_i\}$ provide the desired geometrically transverse spheres for these embedded discs. This will complete the proof of the disc embedding theorem. We will return to this discussion at the end of Chapter 18.

CHAPTER 18

Tower height raising and embedding

Allison N. Miller and Mark Powell

By the end of the previous chapter, we had reduced the proof of the disc embedding theorem to two steps: showing that every union-of-discs-like 1-storey capped tower with at least four surface stages contains mutually disjoint embedded skyscrapers with the same attaching region, and showing that every skyscraper is homeomorphic to the standard 2-handle relative to the attaching region. In this chapter, we will prove the first fact. We will also establish an embedding property for skyscrapers, called the skyscraper embedding theorem, that will be key to the proof of the second fact in Part IV.

Recall that, roughly speaking, a *tower* $\mathcal{T}_n$ with n storeys consists of n capped gropes, each of whose caps are mutually disjoint and have algebraically cancelling double points, where each grope is attached to the collection of thickened accessory circles (that is, the *tip regions* for the caps of the previous grope. A schematic of a tower is shown in Figure 18.1. A *capped tower* with n storeys is obtained from a tower of n storeys by attaching self-plumbed thickened discs to the collection of thickened accessory circles (again, the tip regions) for the top layer of discs in the tower. This new layer of thickened discs, called the *tower caps*, may intersect one another arbitrarily, but their interiors are disjoint from the rest of the capped tower. We also recall the definition of towers and capped towers from earlier for the convenience of the reader (see Section 12.1 for more details). We will often use the notation from the following definition in the proofs in this chapter.

Definition 12.6(3-4) (Towers and capped towers).

- (i) A *tower* $\mathcal{T}_n$ with n storeys is a symmetric generalised tower $G_0 \cup G_1 \cup \cdots \cup G_h$, with only surface and disc stages, such that G_h is a disc stage. Let $0 = i_0 < i_1 < i_2 < \cdots < i_n = h$ be such that the G_{i_j}, for $j = 1, \ldots, n$, are exactly the disc stages.
  - (a) For $j = 0, \ldots, n-1$, the union $G_{i_j+1} \cup G_{i_j+2} \cup \cdots \cup G_{i_{j+1}}$ is called the $(j+1)$th storey.
  - (b) We say that the jth storey has $i_j - i_{j-1} - 1$ surface stages.
  - (c) By definition, the disc stage G_{i_j} is composed of the *jth storey grope caps*. The nth storey grope caps of an n-storey tower are sometimes referred to as the *top storey grope caps*.
- (ii) A *capped tower* $\mathcal{T}_n^c$ with n storeys is a symmetric generalised tower $G_0 \cup G_1 \cup \cdots \cup G_{h+1}$ where $G_0 \cup G_1 \cup \cdots \cup G_h$ is a tower with n storeys and G_{h+1} is a cap stage.

Note that towers are not assumed to be connected, nor is the first stage G_1. So the towers and capped towers in the previous definition are union-of-discs-like, or union-of-spheres-like (capped) towers. In Chapter 17 we found a union-of-discs-like 1-storey capped tower with a collection of geometrically transverse spheres in our ambient 4-manifold. This required either a condition on the fundamental group

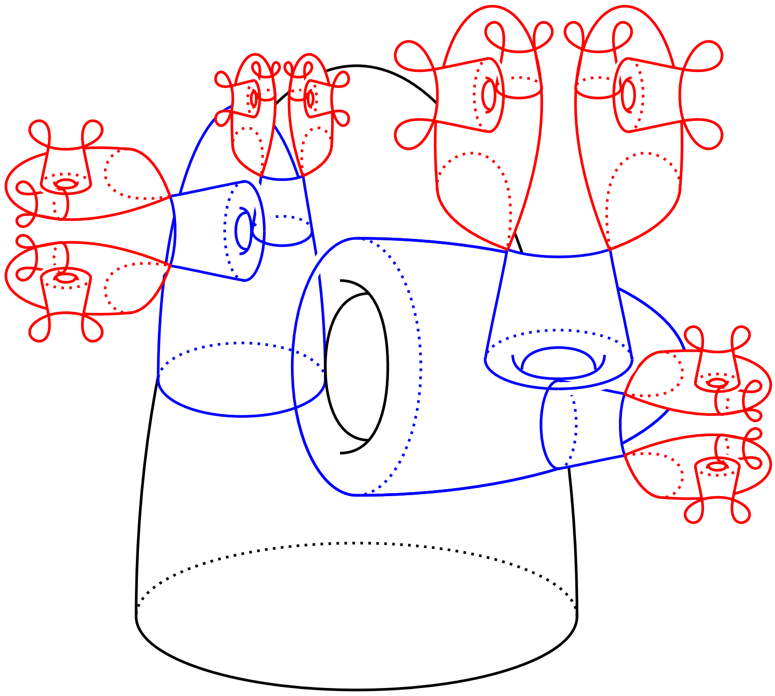


Figure 18.1. A schematic picture of the 2-dimensional spine of a 2-storey tower.

of the ambient manifold or a π_1-nullity condition. In this chapter, we find all the transverse spheres and π_1-nullity necessary for tower building within the capped tower itself, and the ambient 4-manifold will be completely irrelevant.

18.1. The tower building permit

Here is our aim for the near future, following [**FQ90**, Chapter 3].

Proposition 18.1 (Tower building permit). *Let $n \in \mathbb{N}$ and let $\mathcal{T}_n^c$ be a union-of-discs-like or union-of-spheres-like n-storey capped tower, with at least four surface stages in the first storey and at least two surface stages in the nth storey. Then for every $m \in \mathbb{N}$ there exists an $(n+1)$-storey capped tower $\mathcal{T}_{n+1}^c$ contained within $\mathcal{T}_n^c$, with grope height m in the $(n+1)$th storey, such that the first $n-1$ storeys and all the surface stages of the nth storey of $\mathcal{T}_{n+1}^c$ and $\mathcal{T}_n^c$ coincide.*

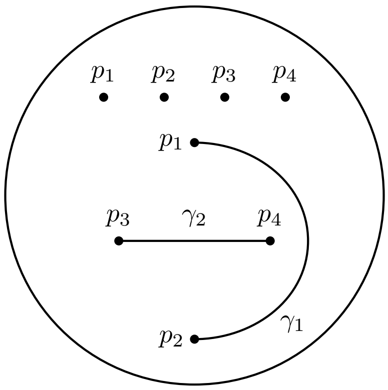
See Definition 17.2 for the precise definition of what it means for the first few stages of two generalised towers $\mathcal{A} \subseteq \mathcal{B}$ to coincide. In particular, it means that the attaching regions of $\mathcal{A}$ and $\mathcal{B}$ are the same.

This tower building permit must have required a fair amount of bribery of city officials since, as we will soon see, it enables us to construct towers with arbitrarily many storeys.

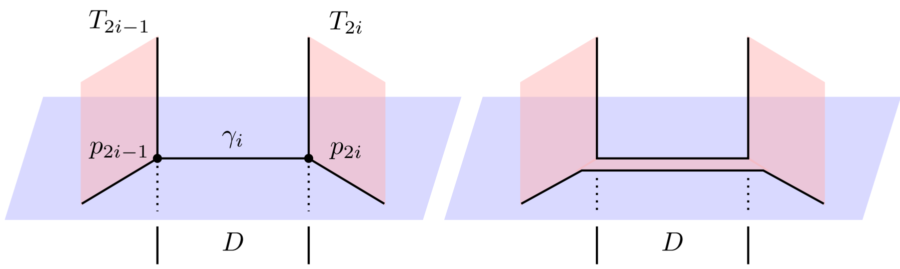
Observe that Proposition 18.1, Lemma 18.2 and Lemma 18.4 below apply to both union-of-discs-like and union-of-spheres-like gropes and towers.

LEMMA 18.2 (Disjoint parallel copies lemma). *Let $G^c(m+1)$ be a union-of-discs-like or union-of-spheres-like capped grope of height $m+1$ for some $m \in \mathbb{N}$. Then there exist arbitrarily many mutually disjoint height m capped gropes within $G^c(m+1)$, the spines of whose bodies are parallel copies of the spine of the first m stages of $G^c(m+1)$. Moreover, all the caps of all the disjoint parallel copies are pairwise disjoint and have algebraically cancelling double points.*

PROOF. Let k be a positive integer. Take k parallel copies of $G^c(m+1)$. The union of this collection is a capped grope of height $m+1$. Note that the caps can intersect one another without restriction, but the body has k times as many disjoint connected components as the bottom surface stage of $G^c(m+1)$. Apply the sequential contraction lemma (Lemma 17.7) to produce a height m capped grope with mutually disjoint caps and algebraically cancelling double points. Since the caps are now mutually disjoint, we have produced k pairwise disjoint capped gropes of height m, such that the spine of the body of each is a parallel push-off of the first m stages of $G^c(m+1)$ as desired. □



(a)



(b) (c)

Figure 18.2. (a) Finding the disjoint arcs $\{\gamma_i\}$. (b-c) Combining two towers bounded by accessory circles to a single tower bounded by a Whitney circle. The latter is formed by taking the boundary connected sum of the two towers along an arc, and then pushing the interior of the new object off the Whitney disc D.

LEMMA 18.3 (Accessory to Whitney lemma). *Let D be a disc stage and let $T_1, \dots, T_{2k}$ be towers attached to the tip regions of D, such that T_{2i-1} and T_{2i} are attached to the tip regions associated to self-plumbings p_{2i-1} and p_{2i} of D respectively, of opposite sign. For every i, there is an arc γ_i on the boundary of D such that the union of γ_i with a parallel copy of itself and the two tip circles (minus* *a small arc of each) forms a Whitney circle for the pair p_{2i-1} and p_{2i}. The union of T_{2i-1}, T_{2i}, and a thickening of a small strip $\gamma_i \times [0,1]$ with interior pushed off the boundary of D, is a tower attached along the Whitney circle and which lies in the complement of D apart from its attaching region.*

Moreover, the arcs $\{\gamma_i\}$ and therefore, the Whitney circles may be chosen to be mutually disjoint, and still disjoint after projecting to the spine of D.

Proof. Choose pairwise disjoint arcs connecting pairs of self-plumbings of opposite signs on the underlying spine discs of D, as shown in Figure 18.2(a). Then take the push-off of γ_i to the boundary of D and a parallel copy of γ_i such that the Whitney move along an embedded Whitney disc with boundary a Whitney circle formed using γ_i, the parallel copy, and the boundary circles of T_{2i-1} and T_{2i}, and with the corresponding framing, would cancel the two double points. Figure 18.2(c) shows the construction of the Whitney circle along with the tower bounded by it in a neighbourhood of the arc γ_i. We are essentially constructing the boundary connected sum of T_{2i-1} and T_{2i} along the arc γ_i. The figure shows a thickened arc on the spine of D which can be pushed into the boundary of D. Push the interior of the strip a bit farther to make it disjoint from D. □

Lemma 18.4. *Every union-of-discs-like or union-of-spheres-like capped tower $\mathcal{T}_n^c$ with at least two surface stages in the first storey, and a decomposition of the second and higher stages into $(+)$- and $(-)$-sides, contains geometrically transverse capped gropes for the second stage surfaces of its first storey satisfying one of the following properties.*

1. *Apart from the transverse point, the body of each transverse capped grope is disjoint from $\mathcal{T}_n^c$, and the caps intersect only the top storey grope caps. Each transverse capped grope is π_1-null in $\mathcal{T}_n^c$. The collection of duals for the $(+)$-side is a mutually disjoint collection, as is the collection of duals for the $(-)$-side.*
2. *Apart from the transverse point, the body of each transverse capped grope is disjoint from $\mathcal{T}_n^c$ and the caps intersect only the tower caps of $\mathcal{T}_n^c$.*

In the first item, the duals for the $(+)$-side and the duals for the $(-)$-side can intersect one another. We will only ever use one of these two collections at a time.

Proof. See Figure 18.3 as well as Remark 12.9 for a division of the second and higher stages of $\mathcal{T}_n^c$ into a $(+)$-side and a $(-)$-side.

We will first construct geometrically transverse gropes for the $(-)$-side using parallel copies of the $(+)$-side and then vice versa. Of course, one may reverse the rôles of $+$ and $-$. We will use the notation of Definition 12.6(3-4) for the stages of $\mathcal{T}_n^c$.

First we prove statement (1). To begin, construct a geometrically transverse capped grope E^1_- for G_2^-, as in the proof of the cap separation lemma (Lemma 17.4), by using two parallel copies of $G_2^+ \cup \cdots \cup G_{i_1}^+$ along with annuli in a neighbourhood of the attaching circle of G_2^+. We obtain a height $i_1 - 2$ union-of-spheres-like capped grope that is geometrically transverse to the $(-)$-side second stage surfaces, and which intersects the first storey grope caps. For each pair of intersections of the first storey grope caps G_{i_1}, there are four pairs of intersections of G_{i_1} with the caps of E^1_-, and four pairs of intersections of caps of E^1_-, as shown in Figure 18.4.

If $n = 1$, stop now. If $n > 1$, there is a disc or surface stage attached to the tips of G_{i_1}. By the accessory to Whitney lemma (Lemma 18.3), we can find Whitney circles bounding subsets of the second storey of $\mathcal{T}_n^c$ along with thin strips. Take eight parallel copies of these Whitney circles and what they bound, corresponding

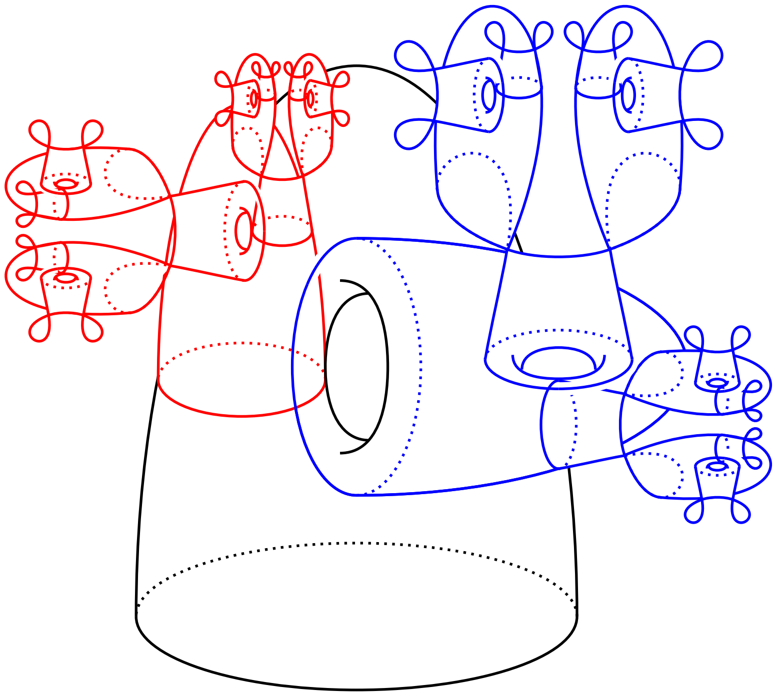


Figure 18.3. Splitting the second and higher stages of a 2-storey tower into $(+)$- and $(-)$-sides. The $(+)$-side is in blue and the $(-)$-side is in red. Only the 2-dimensional spine is shown.

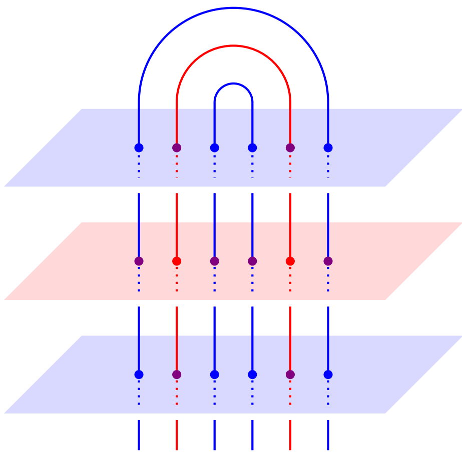
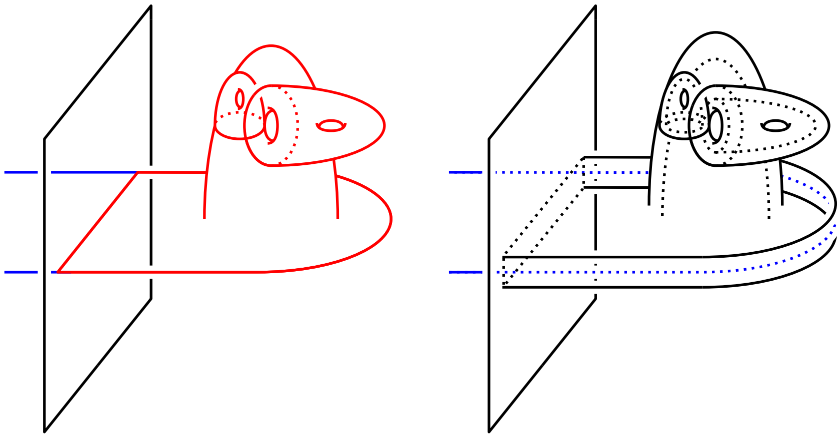


Figure 18.4. A single pair of self-intersections of G_{i_1} (red) becomes four pairs of intersections of G_{i_1} with the caps of E^1_- (purple) and four pairs of intersections of caps of E^1_- (blue).

to the eight pairs of blue and purple intersections shown in Figure 18.4, and perform eight *grope-Whitney moves*. A grope-Whiteny move is just like a Whitney move, except that instead of two parallel copies of a Whitney disc, we may use two parallel copies of a grope (see Figure 18.5). This move replaces each cap of E^1_- with



(a) (b)

Figure 18.5. (a) A grope bounded by a Whitney circle. Only the 2-dimensional spine is shown. (b) The result of the grope-Whitney move. Note that two parallel copies of the grope shown on the right have been used. The blue and black surfaces no longer intersect.

a capped disc-like grope of height $i_2 - i_1 - 1$, which might only be an immersed disc in the degenerate case of $i_2 = i_1 + 1$. Call the resulting union-of-spheres-like capped grope E^2_-. The caps of E^2_- intersect the second storey grope caps on the (+)-side but no longer intersect the first storey grope caps.

Repeat the above procedure up the tower. To achieve option (1), stop at E^n_-, a height $i_n - n - 1$ geometrically transverse union-of-spheres-like capped grope for G^-_2, that, apart from the transverse points, only intersects the top storey grope caps of $\mathcal{T}^c_n$. Note that the capped grope we have constructed lies in the body of $\mathcal{T}^c_n$ (not including the tower caps). Due to the presence of the tower caps, such a neighbourhood, and therefore our geometrically transverse capped gropes, is π_1-null in $\mathcal{T}^c_n$.

Note that the tower $\mathcal{T}^c_n$ has not been modified. Repeat this proof with the rôles of + and − reversed, to produce a union-of-spheres-like capped grope E^n_+ geometrically transverse to G^+_2.

Consider the union-of-spheres-like capped gropes E^n_- and E^n_+. We claim they are each a mutually disjoint union of sphere-like capped gropes. The bodies are mutually disjoint by construction. The claim follows since the caps of $E^n_\pm$ are parallel copies of the grope caps of $\mathcal{T}^c_n$, and these are mutually disjoint. Of course, multiple parallel copies of these grope caps were used in the construction. However, since any given grope cap of $\mathcal{T}^c_n$ corresponds to a unique connected surface in $G^+_2 \cup G^-_2$, the caps of the distinct sphere-like capped gropes constituting $E^n_\pm$ do not intersect one another. Rather, the caps of each constituent sphere-like capped grope intersect one another.

This completes the proof of option (1).

REMARK 18.5. We record here that the constructed geometrically transverse capped gropes E^n_- and E^n_+ are transverse to the $(-)$- and $(+)$-sides of $\mathcal{T}^c_n$ respectively, the capped grope E^n_- intersects only the caps on the $(+)$-side, and E^n_+ intersects only the caps on the $(-)$-side.

To achieve option (2), first construct the transverse capped gropes for option (1) and then go one step further, using the tower caps for a final set of immersed Whitney moves to modify the caps of E^n_- so that the caps are disjoint from the body of the tower, but instead intersect the tower caps. In this case, we only need to perform four Whitney moves for each pair of top storey grope cap intersections of $\mathcal{T}^c_n$ rather than eight as before, but performing more Whitney moves does not hurt us. Call the new transverse grope E^c_- and note that it has no reason to be π_1-null in $\mathcal{T}^c_n$. As before, repeat this process with E^n_+. In this case, since the tower caps intersect arbitrarily, the analogue of Remark 18.5 does not hold, so the caps of E^c_- and E^c_+ may intersect one another, the caps of E^c_- may intersect tower caps on the $(-)$-side, and the caps of E^c_+ may intersect tower caps on the $(+)$-side. ☐

Now we are ready to explain how to increase the number of storeys in a tower by one.

PROOF OF THE TOWER BUILDING PERMIT (PROPOSITION 18.1). Once again, we will use the notation of Definition 12.6(3-4) within the proof. In particular, recall that the i_j are the heights of the disc stages.

We start by constructing geometrically transverse spheres $\{S_i\}$ for the higher stages of the tower. Use Lemma 18.4(2) to find geometrically transverse union-of-spheres-like capped gropes for the second stage surfaces of the first storey that only intersect the tower caps of $\mathcal{T}^c_n$. Contract these until they become immersed spheres $\{S_i\}$ and push the tower caps off the contraction. The set of spheres $\{S_i\}$ is disjoint from $\mathcal{T}^c_n$, apart from the transverse points, by construction. Note that the tower caps of the latter have changed by a regular homotopy, but the body remains unchanged.

Let W' be the complement in $\mathcal{T}^c_n$ of a regular neighbourhood of the spine of the first surface stage of the first storey, so that the $\{S_i\}$ are contained within W'. The second and higher stages of $\mathcal{T}^c_n$ form a union-of-discs-like n-storey capped tower $\mathcal{T}'^c_n$ in W', with geometrically transverse spheres $\{S_i\}$ for the first stage surfaces constructed above, and at least three surface stages in the first storey. Set aside the set $\{S_i\}$ until the end of the proof.

We will now work within the body $\mathcal{T}'_n$ of $\mathcal{T}'^c_n$ which, due to the tower caps, is π_1-null in W'.

Split the second and higher stages of $\mathcal{T}'_n$ into the $(+)$- and $(-)$-side as usual. We are going to add an extra storey to the $(+)$-side. This will create a slightly asymmetric tower, and we will then have to go back and repeat the proof with the $(+)$- and $(-)$-labels switched in order to complete the proof.

Use Lemma 18.4(1) to construct geometrically transverse gropes F^n_+ for the second stage surfaces on the $(+)$-side of $\mathcal{T}'_n$ that only intersect the top storey grope caps for the $(-)$-side (see Remark 18.5), which are moreover π_1-null in $\mathcal{T}'_n$ and thus in W'. Contract the top stage of F^n_+, and push the $(-)$-side top storey grope caps off the contraction (this creates new intersections among the grope caps). By a mild abuse of notation, we still call the result of contraction F^n_+. Observe that F^n_+ is disjoint from $\mathcal{T}'_n$ apart from the transverse points. Moreover, it is still π_1-null in W' since it is a contraction of a π_1-null subset of W'. (Alternatively, it lies inside $\mathcal{T}'_n$ and so is π_1-null in W'.) By construction, the height of F^n_+ is $i_n - n - 2$, which is at least two by hypothesis. Use grope height raising or contraction to arrange for F^n_+ to have height $m+1$, and by a further abuse of notation, still denote the outcome by F^n_+.

In the process of pushing the $(-)$-side top storey grope caps off the contraction of F^n_+, we may have introduced unwanted intersections between these caps. So, raise the height of the top storey of the $(-)$-side by one using grope height raising. This is possible since we have height at least two in the top storey by hypothesis, or, in the case that $n=1$, because we started with height four in the first storey, which in this case is also the top storey). Next, apply the sequential contraction lemma (Lemma 17.7) to arrange once again that the $(-)$-side top storey grope caps are mutually disjoint and have algebraically cancelling double points. Note that these do not currently have tower caps. However, we are still working in $\mathcal{T}'_n$ which is π_1-null in W', so we will soon be able to find tower caps.

By the disjoint parallel copies lemma (Lemma 18.2), there exist arbitrarily many mutually disjoint height m geometrically transverse gropes for the $(+)$-side surfaces of the second stage of the first storey of $\mathcal{T}'_n$, with bodies parallel push-offs of the first m stages of F^n_+ and whose caps are mutually disjoint with algebraically cancelling double points. We call their union $\{\Sigma_+\}$. Since $\{\Sigma_+\}$ lies within $\mathcal{T}'_n$, it is π_1-null in W'.

For each intersection among the $(+)$-side tower caps of $\mathcal{T}'_n$, push the intersections down to the second stage of the first storey, and tube into the mutually disjoint parallel geometrically transverse gropes $\{\Sigma_+\}$. More precisely, for each intersection between tower caps, we obtain 2^{i_n-2} intersections with the second stage $(+)$-side surfaces of $\mathcal{T}'_n$ after pushing down. For each such pushed down intersection, tube into a different parallel copy from the collection $\{\Sigma_+\}$. Thus the $(+)$-side tower caps have been upgraded to capped gropes of height m, forming half of an $(n+1)$th storey for $\mathcal{T}'_n$. That is, we have constructed a generalised tower with $(n+1)$-storeys on the $(+)$-side, n-storeys on the $(-)$-side, but no tower caps. A schematic is shown in Figure 18.6. Moreover, we have done so within $\mathcal{T}'_n$, which is, as noted earlier, π_1-null in W'. Thus we can find tower caps for all the top level discs, on both the $(+)$- and $(-)$-sides. Boundary twist to correct the framing (see Remark 17.10 for specifics), push down any intersections with the body, and tube into parallel copies of the geometrically transverse spheres $\{S_i\}$ constructed in the first step of the proof. This makes the tower caps disjoint from the body, although they may intersect one another arbitrarily.

Now repeat the entire proof with $(+)$- and $(-)$-sides switched in the labelling of the second stages of $\mathcal{T}'_n$ to build a new storey on the $(-)$-side too. This raises the whole tower to have $n+1$ storeys, as desired, and completes the proof of the tower building permit. $\square$

By bootstrapping using the tower building permit, we can upgrade a tower with at least four surface stages in the bottom storey to an arbitrarily tall tower, as follows.

Proposition 18.6 (Tower embedding without squeezing). *Let $\mathcal{T}^c_1$ be a 1-storey union-of-discs-like or union-of-spheres-like capped tower with at least four surface stages in the first storey. For every $m \in \mathbb{N}$, there exists a capped tower $\mathcal{T}^c_m$ with m storeys, embedded within $\mathcal{T}^c_1$, such that the surface stages of the first storey of $\mathcal{T}^c_1$*

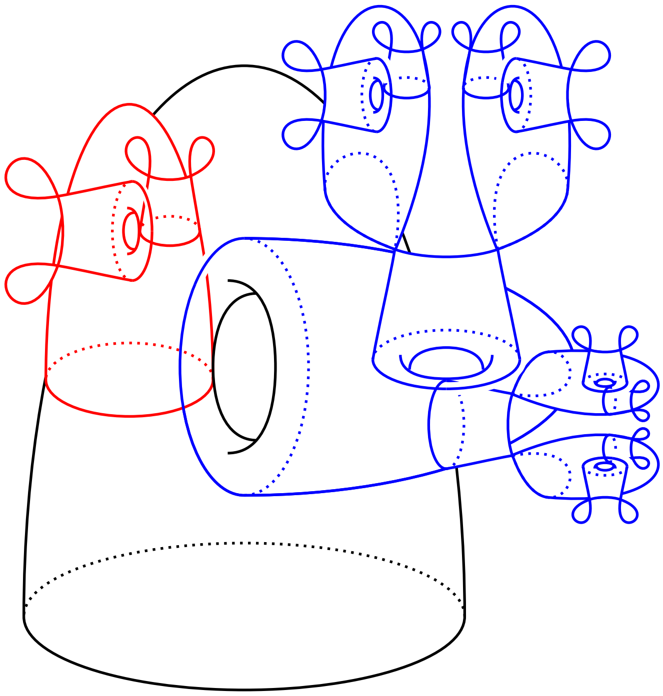


Figure 18.6. A tower with two storeys on the (+)-side (blue) and one storey on the (−)-side (red). Only the 2-dimensional spine is shown.

coincide with the corresponding surface stages of $\mathcal{T}_m^c$*. Moreover, each storey of* $\mathcal{T}_m^c$ *may be arranged to have arbitrarily many surface stages.*

Proof. Apply the tower building permit (Proposition 18.1) $m-1$ times, always to the top storey plus tower caps, ensuring that at each step we produce a top storey with height at least equal to the maximum of two and the height desired for the final sentence. □

18.2. The tower squeezing lemma

Since our ultimate goal is to work with skyscrapers, which are endpoint compactifications of infinite towers, we will need a more powerful version of Proposition 18.6. In particular, given a 1-storey capped tower with at least four surface stages in the first storey, we will find within it a capped tower of arbitrarily many storeys where the higher stages are contained in progressivelly smaller and smaller balls. Consequently, there exists an infinite tower within the original 1-storey capped tower whose closure is homeomorphic to its endpoint compactification. The following lemma shows how to do this.

Since every (finite) tower is a smooth manifold, modulo smoothing corners, it is in particular metrisable. For the rest of this chapter, given a tower, fix a metric it inducing the given (standard) topology.

Lemma 18.7 (Tower squeezing lemma). *Let* $\mathcal{T}_1^c$ *be a* 1*-storey union-of-discs-like or union-of-spheres-like capped tower with at least four surface stages in the first* *storey. Let $\varepsilon > 0$ and let $m \in \mathbb{N}$. Then there is a 2-storey capped tower $\mathcal{T}_2^c \subseteq \mathcal{T}_1^c$ with the same first storey surface stages as $\mathcal{T}_1^c$ and with m surface stages in the second storey. The connected components of the second storey of $\mathcal{T}_2^c$, along with the tower caps, form a collection of mutually disjoint disc-like 1-storey capped towers that lie in mutually disjoint balls, each of radius less than ε.*

As a consequence of the tower squeezing lemma (Lemma 18.7) a union-of-discs-like 1-storey capped tower with at least four surface stages in the first storey contains a *mutually disjoint* union of disc-like 2-storey capped towers.

LEMMA 18.8. *Let $\mathcal{T}^c$ be a union-of-discs-like capped tower with mutually disjoint tower caps.*

(1) The capped tower $\mathcal{T}^c$ is a regular neighbourhood of a 1-complex K, with as many connected components as $\mathcal{T}^c$, and with the property that K becomes a forest, that is a disjoint collection of trees, if the double points of the spine of the tower caps of $\mathcal{T}^c$ are removed.

(2) If $\mathcal{T}^c$ is embedded in a smooth manifold, there is an isotopy of $\mathcal{T}^c$, whose support lies entirely within $\mathcal{T}^c$ and which ends with $\mathcal{T}^c$ lying within an arbitrarily small regular neighbourhood νK of K. Moreover, the path of the isotopy restricted to the attaching region $\partial_-\mathcal{T}^c$ gives an embedding of $\partial_-\mathcal{T}^c \times [0,1]$ within $\mathcal{T}^c$.

PROOF. First we observe that since $\mathcal{T}^c$ has mutually disjoint tower caps, it is indeed a mutually disjoint union of disc-like capped towers, each with mutually disjoint tower caps. Thus we assume henceforth that $\mathcal{T}^c$ is a disc-like capped tower with mutually disjoint tower caps.

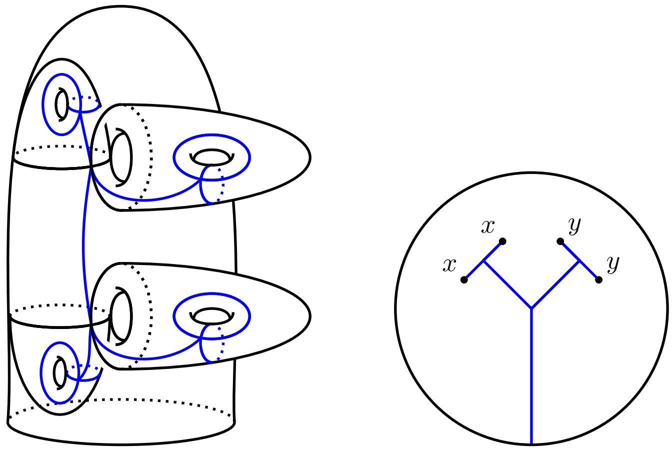
To see that $\mathcal{T}^c$ is a regular neighbourhood of a connected 1-complex, note that, as shown in Chapter 13, every disc-like capped tower has a Kirby diagram consisting of a dotted unlink. In other words, such a capped tower is diffeomorphic to a boundary connected sum of copies of $S^1 \times D^3$, which deformation retracts to a 1-complex. When the caps are mutually disjoint, the loops of the 1-complex correspond to the double point loops of the tower caps. Thus removing the double points leaves a tree. Choices of K for a grope and a disc block are shown in Figure 18.7.

For the second statement, observe that $\mathcal{T}^c$ is built from surface, disc, and cap blocks that only intersect along their attaching and tip regions, since the tower caps are mutually disjoint. The attaching region for each surface, disc, or cap block can be deformed to its tip regions along with some arcs, as shown in Figure 18.8.

The tip regions are the attaching regions for the next set of blocks and can thus be deformed into the next set of tip regions. By continuing this process, we find not only the 1-complex K as desired, but also the path of an isotopy of $\mathcal{T}^c$ to any given regular neighbourhood of K.

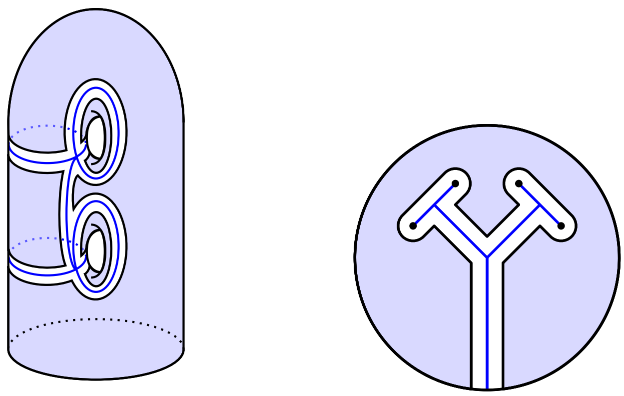
Next, carefully examine the path of $\partial_-\mathcal{T}^c$ under the above isotopy within $\mathcal{T}^c$, as shown in Figures 18.8, 18.9, and 18.10. In the jth stage of $\mathcal{T}^c$, we see 2^{j-1} sheets of $\partial_-\mathcal{T}_1^c$ in this isotopy. Observe that we described an embedding of $\partial_-\mathcal{T}^c \times [0,1]$, as desired. □

PROOF OF THE TOWER SQUEEZING LEMMA (LEMMA 18.7). Let W be the original $\mathcal{T}_1^c$. To begin with, we construct a 2-storey capped tower T_2^c with the same first storey surface stages as $\mathcal{T}_1^c$, with m surface stages in the second storey, and with mutually disjoint tower caps that are π_1-null in the complement in W of the first storey of T_2^c. To achieve this, apply tower embedding without squeezing (Proposition 18.6) to construct a 3-storey capped tower T_3^c in W with m surface stages in the second storey and whose first storey surface stages coincide with those of $\mathcal{T}_1^c$. Discard the tower caps but remember that their existence implies that the third



(a) (b)

Figure 18.7. (a) A choice of K for a height two grope. (b) A choice of K for a disc block, shown in the core D^2 prior to self-plumbing. Points with the same label are identified. Only the 2-dimensional spine is shown in both cases.



(a) (b)

Figure 18.8. An embedding (blue) of the attaching circle times an interval in (a) the case of a surface block and (b) the case of a disc block (compare with Figure 18.7). Only the 2-dimensional spine is shown in both cases.

storey of T_3^c is π_1-null in the complement in W of the first storey of T_3^c. Totally contract the third storey to produce new tower caps for the second storey, and call the result T_2^c. Since the contraction takes place in the third storey, the resulting new tower caps are π_1-null in the complement in W of the first storey of T_3^c, which by construction is also the first storey of T_2^c. Moreover, since the third storey grope caps of T_3^c are mutually disjoint, the tower caps of T_2^c are mutually disjoint.

Let $\widetilde{T}_1^c$ be the tower given by the second storey, along with the tower caps, of T_2^c. Note that this is a union-of-discs-like 1-storey capped tower with mutually disjoint tower caps. Let K be the 1-complex produced by Lemma 18.8 for $\widetilde{T}_1^c$. Choose a

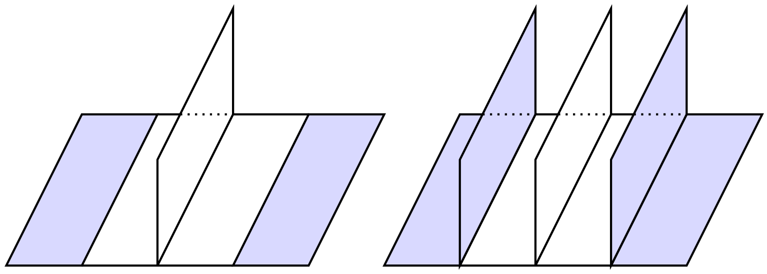
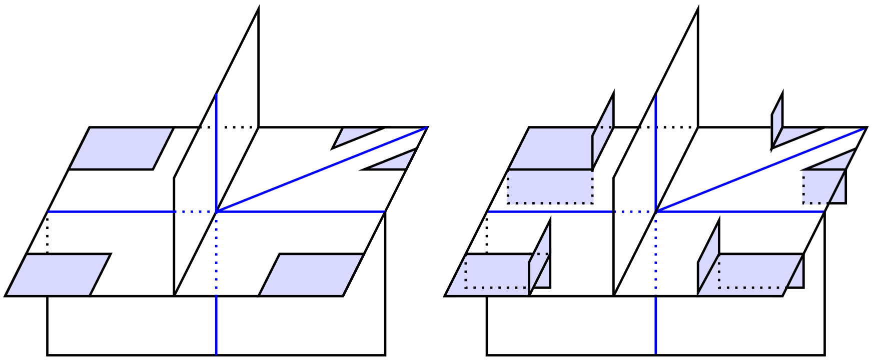


Figure 18.9. A local picture near the attaching circles for the next stage showing the embedding of the attaching circle times an interval in two steps.



(a) (b)

Figure 18.10. A local picture near the attaching circles for the next stage, showing the construction of an embedding of the attaching circle times an interval in two steps.

pairwise disjoint collection of balls in W, each with radius less than ε and disjoint from T_2^c, in one to one correspondence with the connected components of K.

Recall that K is a forest once the double points of the spine of the tower caps of $\widetilde{T}_1^c$ are removed. Since the tower caps are π_1-null in the complement in W of the first storey of T_2^c, so is K. Thus there exists a map $F\colon K \times [0,1] \to W$ whose image is disjoint from the first storey for T_2^c, such that $F|_{K\times\{0\}}\colon K \times \{0\} \to W$ is the original K and such that $F|_{K\times\{1\}}\colon K \times \{1\} \to W$ is an embedding of each of the connected components of K into the mutually disjoint, small balls chosen above.

Perturb the map F so that it is an immersion. That is, henceforth we assume that F is a topological embedding in a neighbourhood of $Y \times [0,1]$ for each neighbourhood Y of a branching point of K, and is a manifold immersion away from the branching points of $K \times [0,1]$. The image $F(K \times [0,1])$ may intersect itself and the capped tower $\widetilde{T}_1^c$, but not the first storey of T_2^c by construction. Push $F(K \times [0,1])$ and $\widetilde{T}_1^c$ in order to move these intersections off $F(K \times [0,1])$, using finger moves along disjoint arcs, and pushing off the $F(K \times \{1\})$ end of $F(K \times [0,1])$. An instance of such a finger move is shown in Figure 18.11.

In the case of self-intersections of $F(K \times [0,1])$, this modifies the map F, and in the case of intersections of $F(K \times [0,1])$ with $\widetilde{T}_1^c$, we isotope the tower. The outcome is an embedding $F'\colon K \times [0,1] \to W$ whose image is disjoint from the capped tower T_2^c apart from at $F(K \times \{0\})$, and such that $F'(K \times \{1\})$ is an embedding

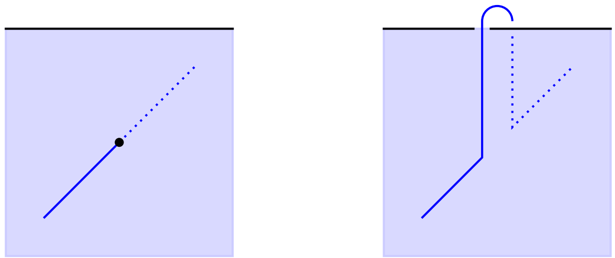


Figure 18.11. Finger move over $F(K \times \{1\})$ to remove an intersection of $F(K \times [0,1])$ with itself, or with $\widetilde{T}_1^c$.

of the connected components of K into mutually disjoint, small balls.

Let νK be a small enough regular neighbourhood of K in W such that the connected components of νK are in one to one correspondence with the connected components of $\widetilde{T}_1^c$. By the construction of K in Lemma 18.8, there is an isotopy H of $\widetilde{T}_1^c$ ending with $\widetilde{T}_1^c$ lying within νK so that the attaching region traces out an embedded $\partial_- \widetilde{T}_1^c \times [0,1]$ in the course of the isotopy H.

Let A be the union of the tip regions of the first storey of T_2^c, which are by definition identified with the attaching region $\partial_- \widetilde{T}_1^c$. Let $A \times [0,1]$ be a collar neighbourhood of A in the first storey of T_2^c, in the sense of manifolds with corners, that is, $\partial A \times [0,1]$ lies in the vertical boundary of the first storey of T_2^c (see Figure 18.12). Let C be the first storey of T_2^c minus $A \times [0,1]$. We define an isotopy of the tower T_2^c, during which C remains fixed. On $\widetilde{T}_1^c$, this consists just of the isotopy H from before. All points in C remain fixed while the collar $A \times [0,1]$ gets stretched out along the embedding of $\partial_- \widetilde{T}_1^c \times [0,1]$. Since we are using an embedded copy of $\partial_- \widetilde{T}_1^c \times [0,1]$, the final result is an isotoped copy of T_2^c, with no new self-intersections, such that the entire spine of $\widetilde{T}_1^c$ lies in νK.

Finally, isotope the connected components of νK to mutually disjoint balls by extending the isotopy of K determined by $F'(K \times [0,1])$. This moves all of $\widetilde{T}_1^c$ into mutually disjoint small balls, as required, while stretching out $A \times [0,1]$ still further. A schematic for the last two steps is shown in Figure 18.12. □

18.3. The tower and skyscraper embedding theorems

The work of this chapter culminates in the following theorem.

THEOREM 18.9 (Tower embedding theorem). *Let $\mathcal{T}_1^c$ be a union-of-discs-like or union-of-spheres-like 1-storey capped tower with at least four surface stages in the first storey. Within $\mathcal{T}_1^c$ there exists an infinite compactified tower $\widehat{\mathcal{T}}_\infty$ with the same first storey surface stages as $\mathcal{T}_1^c$ such that the following holds.*

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

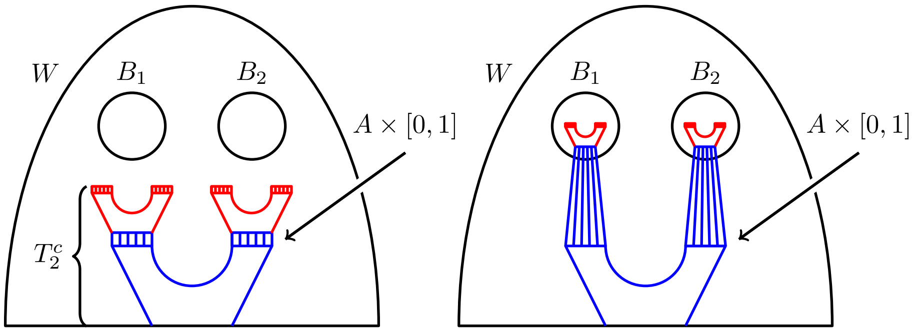


Figure 18.12. Moving the second storey and tower caps of T_2^c, namely $\widetilde{T}_1^c$, into mutually disjoint small balls. The capped tower $\widetilde{T}_1^c$ is shown in red. Note that the first story of T_2^c, shown in blue, remains unchanged, except that a collar of the tip region gets stretched out.

As we will see in the proof of the tower embedding theorem (Theorem 18.9), we need the tower squeezing lemma (Lemma 18.7) in order to obtain an embedding of the endpoint compactification $\widehat{\mathcal{T}}_\infty$ rather than just an infinite tower. However, the third condition in the theorem is also needed in Part IV beyond this, and may require further additional squeezing. However, by choosing small enough balls early on in the squeezing process, the third condition can easily be arranged.

Proof. Iteratively apply the tower squeezing lemma (Lemma 18.7) to $\mathcal{T}_1^c$, with $\varepsilon_j = 1/j$ and $m_j = 2^{j+1}$ on the jth iteration. Each time apply the lemma to the top storey and tower caps of each tower. Squeeze further if necessary to arrange the third condition. This shows that $\mathcal{T}_1^c$ contains an infinite tower $\mathcal{T}_\infty$ with the same first storey surface stages and whose higher stages are contained in progressively smaller, mutually disjoint balls, satisfying the three given conditions. Define $\widehat{\mathcal{T}}_\infty$, (with suggestive notation) to be the closure of $\mathcal{T}_\infty$ in $\mathcal{T}_1^c$. As a closed subset of the compact space $\mathcal{T}_1^c$, the space $\widehat{\mathcal{T}}_\infty$ is compact. We will show that indeed $\widehat{\mathcal{T}}_\infty$ is the endpoint compactification of $\mathcal{T}_\infty$.

We need to show that the limit points added to $\mathcal{T}_\infty$ to obtain $\widehat{\mathcal{T}}_\infty$ can be identified with the ends of $\mathcal{T}_\infty$ and that the subspace topology on $\widehat{\mathcal{T}}_\infty$ is the topology on an endpoint compactification. Let x be a point in $\widehat{\mathcal{T}}_\infty \smallsetminus \mathcal{T}_\infty$. Then x is a limit of a convergent sequence of points in $\mathcal{T}_\infty$, call it $\{a_j\}$. Each a_j is contained in some finite truncation $\mathcal{T}_\infty^{\leq k_j}$ of $\mathcal{T}_\infty$. Since x is not in $\mathcal{T}_\infty$, the sequence $\{k_j\}$ must be unbounded: this follows from the fact that each finite truncation of $\mathcal{T}_\infty$ is a compact subset of the Hausdorff space $\widehat{\mathcal{T}}_\infty$ and is consequently closed and contains all its limit points. In other words, the points a_j must eventually leave every finite truncation of $\mathcal{T}_\infty$. Let $\{a_{j_\ell}\}$ denote a subsequence of $\{a_j\}$ such that $\{k_{j_\ell}\}$ is strictly increasing. Note that $\{a_{j_\ell}\}$ also converges to x. Let U_{j_ℓ} be the component of the complement of $\mathcal{T}_\infty^{\leq k_{j_\ell}}$ in $\mathcal{T}_\infty$ containing a_{j_ℓ}. Then we associate to x the sequence $(U_{j_1}, U_{j_2}, \dots)$. By the definition of ends of a space, such a sequence corresponds to a unique end of $\mathcal{T}_\infty$ (the sequence $(U_{j_1}, U_{j_2}, \dots)$ might not be an end of $\mathcal{T}_\infty$ on the nose; in general we need to add some intervening elements to obtain an end. However, the resulting end is uniquely determined.)

We need to show that the above correspondence between the set of limit points of $\widehat{\mathcal{T}}_\infty$ contained in $\widehat{\mathcal{T}}_\infty \smallsetminus \mathcal{T}_\infty$ and the set of ends of $\mathcal{T}_\infty$ is a bijection. Certainly any end of $\mathcal{T}_\infty$ gives rise to a sequence of points, obtained by successively choosing points in components of complements of progressively larger finite truncations of $\mathcal{T}_\infty$. Since these components lie in metric balls of progressively smaller radii converging to zero, by construction, the sequence converges. Since the sequence of points leaves every finite truncation of $\mathcal{T}_\infty$, the limit point must lie in $\widehat{\mathcal{T}}_\infty \smallsetminus \mathcal{T}_\infty$. Moreover, by the same argument, any two sequences corresponding to the same end have the same limit point. Additionally, in the construction, connected components of higher levels of $\mathcal{T}_\infty$ were placed in small balls with positive distance between them. Consequently, sequences of points associated with distinct ends of $\mathcal{T}_\infty$ cannot have the same limit point either. This finishes the proof that the set of endpoints of $\mathcal{T}_\infty$ is in bijective correspondence with the points of $\widehat{\mathcal{T}}_\infty \smallsetminus \mathcal{T}_\infty$.

It is easy to see that the topology on the endpoint compactification coincides with the subspace topology by inspecting the definition of both. By Definition 12.17, the endpoint compactification topology on $\widehat{\mathcal{T}}_\infty$ is generated by open sets of $\mathcal{T}_\infty$ and sets $V \subseteq \widehat{\mathcal{T}}_\infty$ such that $V \cap \mathcal{T}_\infty$ is a component U of the complement of some finite truncation $\mathcal{T}_\infty^{\leq k}$ of $\mathcal{T}_\infty$, and V contains the endpoints of $\mathcal{T}_\infty$ associated with U. By construction, all such sets are also open in the subspace topology. Similarly, any open set in the subspace topology is open in the endpoint compactification topology. $\square$

Recall the following definition.

Definition 12.21 (Skyscrapers and open skyscrapers). A disc-like infinite tower $\mathcal{S}$ is said to be an *open skyscraper* if it satisfies the following two conditions.

(a) (Replicable) Each storey has at least four surface stages.
(b) (Boundary shrinkable) $\sum_{i=1}^{\infty} N_j/2^j$ diverges, where N_j is the number of surface stages in the jth storey.

The union of the $2i-1$ and the $2i$ storeys of an open skyscraper is called its ith *level*.

The endpoint compactification of an open skyscraper is called a *skyscraper*, denoted by $\widehat{\mathcal{S}}$. Similarly, given a skyscraper $\widehat{\mathcal{S}}$, the corresponding open skyscraper is denoted by $\mathcal{S}$.

In this chapter, we have established the following theorem.

Theorem 18.10 (Skyscraper embedding theorem). *Let $\widehat{\mathcal{S}}$ be a skyscraper. Let $\mathcal{T}_2$ be a connected component of some level of $\widehat{\mathcal{S}}$. Then there is an embedding of a skyscraper $\widehat{\mathcal{S}}' \subseteq \mathcal{T}_2$ such that the attaching region $\partial_-\widehat{\mathcal{S}}'$ and the first surface stage of $\widehat{\mathcal{S}}'$ agree with those of $\mathcal{T}_2$, and the connected components of the levels of $\widehat{\mathcal{S}}'$ with indices greater than or equal to two lie in arbitrarily small, mutually disjoint balls.*

Proof. Each level of $\widehat{\mathcal{S}}$ is a mutually disjoint collection of disc-like 2-storey towers. For any such disc-like 2-storey tower $\mathcal{T}_2$, totally contract the second storey to get tower caps for the first storey, resulting in a 1-storey capped tower $\mathcal{T}_1^c$. Apply the tower embedding theorem (Theorem 18.9) to produce an infinite compactified tower $\widehat{\mathcal{T}}_\infty =: \widehat{\mathcal{S}}'$. Note that $\mathcal{T}_2$ (and thus $\mathcal{T}_1^c$) and $\widehat{\mathcal{T}}_\infty$ have the same first storey surface stages. $\square$

Remark 18.11. The tower embedding theorem in particular implies that a union-of-discs-like 1-storey capped tower $\mathcal{T}_1^c$ contains mutually disjoint embedded skyscrapers with the same attaching region as $\mathcal{T}_1^c$. As stated at the start of this chapter, this reduces the proof of the disc embedding theorem to showing that every skyscraper contains a flat embedded disc whose framed boundary coincides with the attaching region of the skyscraper. In Part IV, we will show that every skyscraper is homeomorphic, relative to the attaching region, to the standard 2-handle, completing the proof of the disc embedding theorem. In particular, this will also show that every union-of-discs-like 1-storey capped tower $\mathcal{T}_1^c$ contains mutually disjoint, embedded flat discs whose framed boundaries coincide with the attaching region of $\mathcal{T}_1^c$.

18.4. Proof of the disc embedding theorem assuming Part IV

The tower embedding theorem combined with Proposition 17.12 from the previous chapter gives rise to the following result.

Proposition 18.12. *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 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 an immersed collection

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

of discs in M such that the following holds.

1. *For each i, the discs f_i and f_i' have the same framed boundary.*
2. *There is an immersed collection*
   $$\overline{G} = (\overline{g}_1, \dots, \overline{g}_n)\colon S^2 \sqcup \cdots \sqcup S^2 \looparrowright M$$
   of framed geometrically transverse spheres for F' in M, such that for every i, the sphere $\overline{g}_i$ is homotopic to g_i.
3. *The intersections and self-intersections among F' are algebraically cancelling.*
4. *Whitney circles for the intersections and self-intersections within F' bound a collection of mutually disjoint skyscrapers, that apart from their attaching regions lie in the complement in M of both F' and $\overline{G}$, and such that the framing of the attaching region agrees with the Whitney framing.*

Proof. Apply Proposition 17.12 to obtain the conclusion with a union-of-discs-like 1-storey capped tower in place of the skyscrapers. By the tower embedding theorem (Theorem 18.9), the union-of-discs-like 1-storey capped tower contains within it a collection of mutually disjoint skyscrapers, with the same framed attaching regions. This completes the proof of the proposition. □

In Part IV we will prove that every skyscraper is homeomorphic to a 2-handle relative to its attaching region.

Theorem 27.1. *For every skyscraper $\widehat{\mathcal{S}}$ there is a homeomorphism of pairs*

$$(\digamma, \partial\digamma)\colon \big(\widehat{\mathcal{S}}, \partial_-\widehat{\mathcal{S}}\big) \cong \big(D^2 \times D^2, S^1 \times D^2\big)$$

that is a diffeomorphism on a collar of $\partial_-\widehat{\mathcal{S}}$, such that, if $\Phi\colon S^1 \times D^2 \to \partial_-\widehat{\mathcal{S}}$ is the attaching region, then $\partial\digamma \circ \Phi = \mathrm{Id}_{S^1 \times D^2}$.

Assuming Theorem 27.1, we complete the proof of the disc embedding theorem.

Proof of the disc embedding theorem. Proposition 18.12 starts with the hypotheses of the disc embedding theorem. Apply Theorem 27.1 to replace the mutually disjoint skyscrapers in the conclusion by a disjointly embedded collection of 2-handles $D^2 \times D^2$, with the attaching circles $S^1 \times \{0\}$ mapping to the Whitney circles. The cores $D^2 \times \{0\}$ then give rise to a flat embedding of a mutually disjoint collection of Whitney discs $\{V_k\}$ for the intersections and self-intersections among the $\{f'_i\}$. Perform the Whitney move using the Whitney discs $\{V_k\}$ to obtain the desired flat embedding

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

with $\overline{f}_i$ regularly homotopic to f'_i for every i. In particular, for every i the discs $\overline{f}_i$ and f'_i, and therefore the discs $\overline{f}_i$ and f_i, have the same framed boundary. Since the Whitney discs $\{V_k\}$ are disjoint from the framed, geometrically transverse spheres

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

these spheres are still geometrically transverse to the discs $\{\overline{f}_i\}$ obtained from the Whitney move. We already had that $\overline{g}_i$ is homotopic to g_i in the conclusion of Proposition 18.12, so this remains true. This completes the proof of the disc embedding theorem, modulo Theorem 27.1. $\square$

Part III

Interlude

In this part, we take a break from the proof of the disc embedding theorem. In Chapter 19 we describe *good groups* in greater detail, proving that all elementary amenable groups are good. In Chapter 20 we show how to use the disc embedding theorem to prove the 5-dimensional *s-cobordism theorem* with good fundamental groups and smooth input, and the *Poincaré conjecture* for smooth 4-dimensional homotopy spheres. We also prove the *sphere embedding theorem*.

In Chapter 21 we present the flowchart in Figure 21.1, explaining the logical dependence among key statements in the development of topological 4-manifold theory. This describes, for example, the input needed to establish fundamental tools such as topological transversality, the immersion lemma, and the existence of normal bundles for locally flat submanifolds in dimension four. We outline multiple proofs of landmark results in the theory, such as the classification of closed, simply connected, topological 4-manifolds up to homeomorphism, the Poincaré conjecture for topological homotopy 4-spheres, and the existence of exotic smooth structures on $\mathbb{R}^4$.

In the two remaining chapters in this part, we assume additional ingredients from Chapter 21 not proved in this book. Chapter 22 is an introduction to surgery theory for 4-manifold topologists. We show, in particular, how to use the category preserving sphere embedding theorem to prove the exactness of the surgery sequence in the topological category for good fundamental groups. We then sketch the application of the surgery sequence to the classification of closed, simply connected, topological 4-manifolds up to homeomorphism. Finally in Chapter 23 we discuss and relate several open problems and conjectures in the area.

CHAPTER 19

Good groups

Min Hoon Kim, Patrick Orson, JungHwan Park, and Arunima Ray

The disc embedding theorem is only known to hold in ambient 4-manifolds with *good* fundamental group. Consequently, many of Freedman and Quinn's topological results, such as the exactness of the surgery sequence in dimension four and the 5-dimensional s-cobordism theorem, are restricted by a hypothesis on good groups. The determination of exactly which groups are good is one of the most important open problems in 4-manifold topology.

Remarkably, the good group hypothesis is only used in a single step of the proof of the disc embedding theorem, namely in Proposition 17.8. The definition of a good group, which we soon recall, reflects precisely what is needed in this step of the proof. In this chapter, we explore what is known about good groups.

First we define a *double point loop* for a capped grope. Choose a base point in the attaching circle of the grope. We obtain a double point loop for each intersection between caps C_1 and C_2 of the grope, where possibly $C_1 = C_2$ (recall that cap intersections are the only allowed intersections within a grope). The double point loop associated to such an intersection point starts at the base point, travels up the grope to the attaching circle of C_1, travels to the double point, switches to the other sheet, namely C_2, then travels to the attaching circle of C_2, and then back down the grope to the base point. A double point loop is only allowed to change sheets at a single, prescribed double point. The fundamental group of a capped grope is freely generated by the double point loops of its cap intersections. This can be seen, for example, from the Kirby diagram for a capped grope (see Chapter 13). Switching the order of C_1 and C_2 replaces the fundamental group element associated with a double point loop by its inverse. Additionally, the fundamental group element determined by a double point loop does not depend on the basing path, since the body of a capped grope is π_1-null in the capped grope. Thus, we will generally refer to *the* double point loop for a cap intersection of a capped grope, with the understanding that we may change the orientation or basing path. Of course, we may define double point loops for an immersed disc in the same manner, by thinking of an immersed disc as a (degenerate) disc-like capped grope of height zero, as on page 158.

Definition 12.12. A group Γ is said to be *good* if for every height 1.5 disc-like capped grope G, with some choice of basepoint, and for every group homomorphism $\phi\colon \pi_1(G) \to \Gamma$, there exists an immersed disc $D \looparrowright G$ whose framed boundary coincides with the attaching region of G, such that the double point loops of D, considered as fundamental group elements by making some choice of basing path, are mapped to the identity element of Γ by ϕ.

Since every homomorphism to the trivial group is trivial and every capped grope contains an immersed disc whose framed boundary coincides with the attaching region, produced by contraction, the trivial group is good. In [**Fre84**, pp. 658–659] (see also [**FQ90**, Section 5.1]), Freedman showed that the infinite cyclic group, as well as any finite group, is good. We give the proofs below. The proof will depend on a careful analysis of how double point loops evolve in the grope height raising (Proposition 17.3) and the contraction and push off (Section 15.2.5) operations. First we consider the contraction and push off operations in the next lemma. A pair of caps C_1 and C_2 attached to dual curve on a surface, so that $\partial C_1 \cap \partial C_2$ is a single point, are called *dual caps.*

LEMMA 19.1. *Let $G^c(n)$ be a capped grope of height n. Let C_1 and C_2 be dual caps. Let c_1 and c_2 be two caps of $G^c(n)$ such that $C_1 \cap c_1$ and $C_2 \cap c_2$ are both nonempty. We allow $c_1 = c_2$ and for either of the c_i to be (a parallel copy of) C_1 or C_2.*

After contracting the grope along (push-offs of) the caps C_1 and C_2 and pushing c_1 and c_2 off the contraction, each pair of intersection points (p_1, p_2), with p_1 an intersection point between C_1 and c_1, and p_2 an intersection point between C_2 and c_2, gives rise to two new intersection points between c_1 and c_2. For $i = 1, 2$, let γ_i be the double point loop for the intersection point p_i between C_i and c_i. The double point loop associated to both of the two new corresponding intersection points is $\gamma_1 \cdot \gamma_2^{-1}$.

PROOF. The proof consists of Figure 19.1, which shows the relationship of the new double point loops with the old double point loops. □

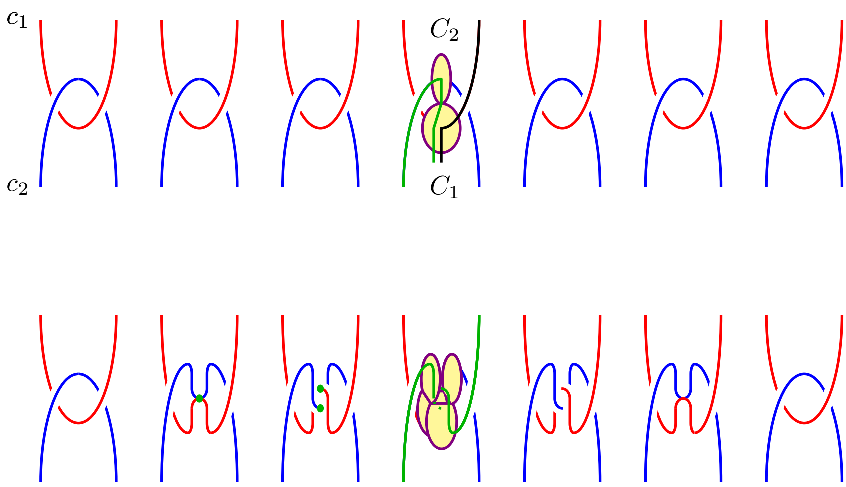


Figure 19.1. New double point loops after a contraction followed by push off. We do not draw the surface being contracted in the figure. Compare with Figure 15.8.
Top: Dual caps C_1 and C_2 are shown in yellow. The caps c_1 and c_2 are shown in red and blue respectively. Black and green show pieces of the double point loops γ_1 and γ_2 for the intersection between C_1 and c_1, and between C_2 and c_2 respectively.
Bottom: Green shows a piece of a new double point loop. Most of the piece shown is in the middle panel, but there is a (trivial) finger extending backwards in time, with the peak at the new double point. A similar extension to the right, which we do not show, would give the double point loop for the second new intersection point, which is visible in the second image from the right.

THEOREM 19.2. *Every finite group is good.*

PROOF. We already know that the trivial group is good. Let Γ be a finite group of order $n+1$, for some $n \geq 0$, and let G be a height 1.5 disc-like capped grope. Fix a basepoint in the attaching region and a homomorphism $\phi\colon \pi_1(G) \to \Gamma$. Use grope height raising (Proposition 17.3) to construct a grope $G' \hookrightarrow G$, with the same attaching region and height n. The map ϕ induces a map $\phi\colon \pi_1(G') \to \Gamma$ since $G' \subseteq G$ and the attaching regions coincide. Fix a bijection between the n surface stages of G' and the nontrivial elements of Γ. In other words, enumerate the nontrivial elements of Γ as $\{\gamma_1, \gamma_2, \dots, \gamma_n\}$. At each intersection among the caps of G', choose a first and second sheet. This fixes the images of the double point loops in Γ, by requiring double point loops to go from the first to the second sheet.

Now contract the top stage of G'. This creates some new intersections, when we double the caps. Since parallel copies of caps are used, the new double point loops are parallel copies of the previous ones, by appropriate choice of basing paths. Assign first and second sheets at these new intersections by letting the first sheet be the parallel of the sheet previously designated as first. Push off caps at any intersections whose double point loops were mapped by ϕ to γ_n. Make sure to push the first sheet off the second. The new grope still has a map from its fundamental group to Γ. By Lemma 19.1, the new intersections created by pushing off in this step have double point loops given by $\gamma_n \cdot \gamma_n^{-1}$, that is, they are mapped to the trivial element in Γ.

Now perform this process iteratively: at the ith step, contract the $(n+1-i)$th stage of the capped grope, assign first and second sheets as before, then push off intersections mapped by $\phi\colon \pi_1(G') \to \Gamma$ to γ_{n+1-i}. After contracting the n stages of G', we are left with an immersed disc D whose framed boundary coincides with the attaching region of G' and thus G, and such that all the double point loops are mapped to the identity element of Γ. This completes the proof. □

Before proving that the infinite cyclic group is good, we investigate the behaviour of double point loops under the operation of grope height raising.

LEMMA 19.3. *Let G be a height 1.5 disc-like capped grope. For any $i \geq 0$, there exists a height 2^i capped grope G', contained within G with the same attaching region, such that each double point loop of G' has length at most 7^{2i} in the double point loops of G.*

PROOF. We recall the proof of grope height raising (Proposition 17.3). It alternates raising the height of the $(+)$- and $(-)$-sides of G. Start with the $(-)$-side, which consists only of caps at the beginning. First, we use the cap separation lemma (Lemma 17.4) to the make the $(+)$-side and $(-)$-side caps disjoint. This changes the $(-)$-side caps of G by a push off operation. By Lemma 19.1 each new intersection point has a double point loop given by the product of two double point loops of previous intersections.

The $(+)$-side caps also change during cap separation. As shown in Figure 19.2, each double point loop is changed by conjugation by a double point loop.

Then we produce geometrically transverse union-of-spheres-like gropes for the $(-)$-side by using parallel copies of the $(+)$-side grope, which has height one at the beginning. Next tube the intersections among the $(-)$-side caps into this grope (after first pushing down to the bottom stage of the $(-)$-side if necessary). The $(-)$-side caps now become capped gropes, which are still on the $(-)$-side by definition. Since parallel copies of the $(+)$-side caps were used in this process, the new $(-)$-side caps now intersect the $(+)$-side caps. From Figure 19.3, we see that a double point loop for a new intersection between a $(+)$-side cap and a $(-)$-side cap goes through two old intersections; these have length five in terms of double point loops of the

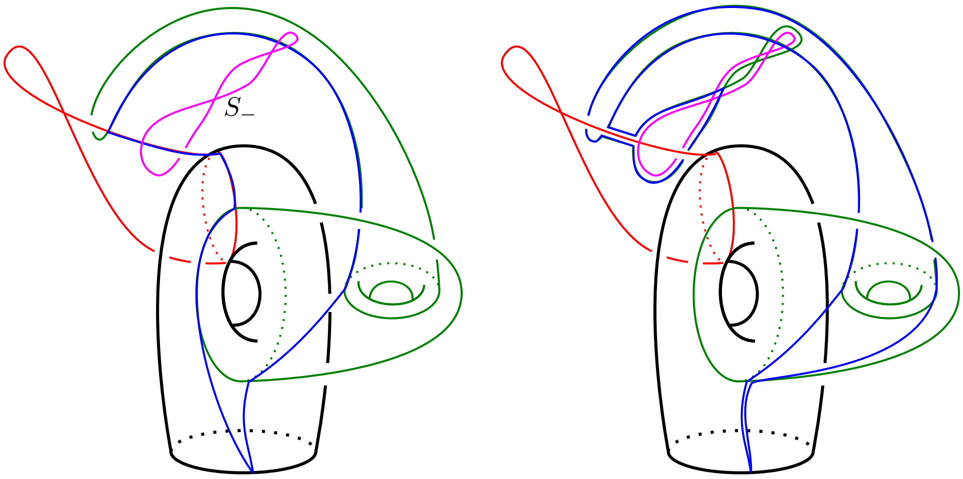


Figure 19.2. (a) We see a disc-like capped grope of height 1.5, after the first step of the cap separation lemma. More precisely, parallel copies of the (+)-side (green) have been used to produce a sphere S_- (pink), geometrically transverse to the (−)-side (red). We only show a single (+)-side cap. The (+)-side cap intersects the (−)-side cap. The double point loop for the intersection point is shown in blue.
(b) The intersection point between the (+)- and (−)-side cap has been removed by tubing into S_-. The double point loop for the new (+)-side intersection point is shown in blue. Note that it passes through an intersection of S_- (arising from an intersection point of the (+)-side), as well as the old intersection point between the (−)- and (+)-side caps. More precisely, the double point loop is a conjugate of a double point loop for a (+)-side intersection point by a double point loop for an intersection between (−)- and (+)-side caps.

original grope. Similarly, a double point loop for a new intersection between two (−)-side caps goes through three old intersections, two of which are produced by contraction and push off operations; these have length seven in terms of double point loops of the original grope.

Putting all of the previous steps together, we see that each of the final double point loops can be written as a product of at most seven double point loops of the original grope.

We repeat this process, switching between (+)- and (−)-sides. At each step we apply the cap separation lemma and then tube intersections into a union-of-spheres-like geometrically transverse grope constructed using parallel copies. The process produces asymmetric gropes which may be contracted to be made symmetric. Contractions do produce new double point loops, but since they are parallel push-offs of the previous double point loops, we may ignore them in our analysis.

From the recurrence relations (17.1) stated in the proof of grope height raising, we know that in $2i$ steps of grope height raising, followed by a final step of contraction, we can produce a grope of height equal to F_{2i+2}, where $\{F_k\}$ is the Fibonacci sequence with $F_0 = F_1 = 1$.

We claim that in $2i$ steps, G can be promoted to a capped grope G' of height at least 2^i, with the same attaching region. That is, we need to show that $F_{2i+2} \geq 2^i$. The claim follows from an easy proof by induction. The base case is trivial. The inductive case uses that $F_{2i+2} = F_{2i} + F_{2i+1} \geq 2F_{2i}$.

Thus the grope G' produced by $2i$ steps of grope height raising has at least 2^i surface stages, and its double point loops can be written as words of length at most 7^{2i} in the double point loops of the original grope G. Perform a contraction operation to ensure that G' has height precisely 2^i. □

THEOREM 19.4. *The infinite cyclic group $\mathbb{Z}$ is good.*

PROOF. Let G be a height 1.5 disc-like capped grope, with a homomorphism $\phi\colon \pi_1(G) \to \mathbb{Z}$. The grope G has a finite number of double point loops. Let N be an upper bound for the absolute values of the image in $\mathbb{Z}$ under ϕ of these double point loops. Use grope height raising (and a final contraction step), to produce a grope G', contained within G with the same attaching region, with 2^i surface stages and whose double point loops are length at most 7^{2i} in the double point loops of G by Lemma 19.3. Thus, the image of the double point loops of G' in $\mathbb{Z}$ under ϕ is bounded above by $N' := 7^{2i}N$.

As in the proof of Theorem 19.2, choose first and second sheets for all cap intersections of G'. Contract the top stage of G', and push off all intersections (pushing the first sheet off the second) whose double point loops map between $N'/2$ and N' in $\mathbb{Z}$. Assign first and second sheets to new intersections as previously. Then contract the top stage again, and this time push off all intersections (the first sheet over the second) whose double point loops map between $-N'$ and $-N'/2$ in $\mathbb{Z}$. By Lemma 19.1, the double point loops for the resulting grope, which has two fewer surface stages than G', have image in $\mathbb{Z}$ bounded by $-N'/2$ and $N'/2$. Repeat this process to contract the top 2^i stages of G' to produce a capped grope (or possibly an immersed disc) such that the image of the collection of double point loops lies in the interval

$$\left[-\frac{N'}{2^{2^{i-1}}}, \frac{N'}{2^{2^{i-1}}}\right].$$

In our case, we have that $N' = 7^{2i}N$, where N is fixed. Choose i so that $\frac{7^{2i}}{2^{2^{i-1}}} \cdot N < 1$. This is possible since $\lim_{i\to\infty} \frac{7^{2i}}{2^{2^{i-1}}} = 0$. Thus for such a choice of i, possibly after a few final contraction operations, we produce an immersed disc D with the same attaching region as G whose double point loops are mapped to $0 \in \mathbb{Z}$ by ϕ. □

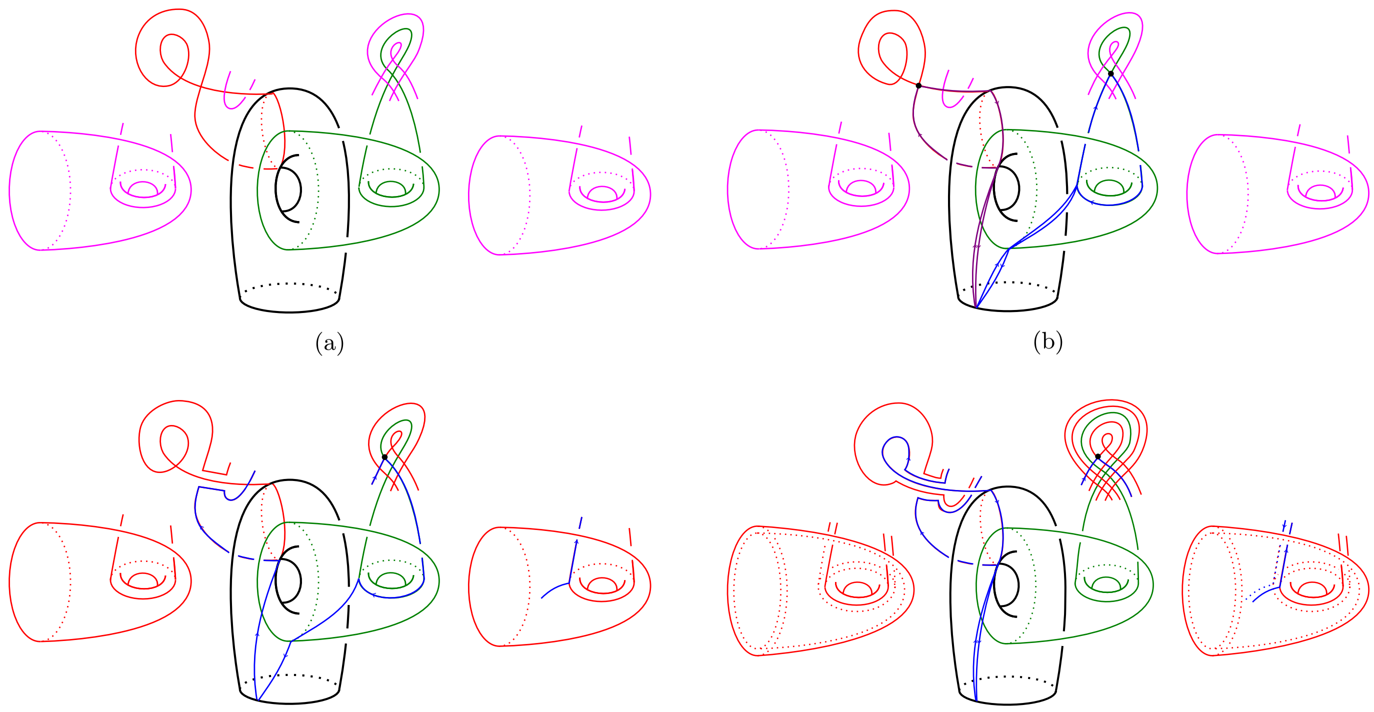
We also know the following properties of good groups.

PROPOSITION 19.5 ([**Fre84**, p. 660; **FQ90**, Exercise 2.9; **FT95a**, Lemma 1.2]). *The class of good groups is closed under*

1. *subgroups,*
2. *quotients,*
3. *extensions,*
4. *colimits (or direct limits) and*
5. *passage to a larger group in which the original is a finite index subgroup.*

PROOF. Let G be a disc-like capped grope of height 1.5. The group $\pi_1(G)$ is a free group generated by the double point loops $a_1, \ldots, a_k$ for the cap intersections of G.

For (1), let Γ be a good group and let $\Gamma' \leq \Gamma$ be a subgroup. Suppose that $\phi\colon \pi_1(G) \to \Gamma'$ is a homomorphism. Compose with the inclusion $\iota\colon \Gamma' \hookrightarrow \Gamma$ to get the map $\iota \circ \phi\colon \pi_1(G) \to \Gamma$. Since Γ is good, we obtain a disc $D \looparrowright G$ such that the double point loops are mapped by $\iota \circ \phi$ to 1_Γ. Since the map ι is of course injective,



(a) (b) (c) (d)

Figure 19.3. (a) A disc-like capped grope of height 1.5 after cap separation, divided into a (−)-side (red) and a (+)-side (green). The transverse sphere-like grope (pink) for the (−)-side cap shows only a piece of the annulus joining the two parallel copies of the (+)-side. (b) Double point loops (purple and blue) for a (−)-side and (+)-side cap intersection. (c) After tubing a (−)-side cap intersection into the transverse grope, a new intersection between the new (−)-side cap and a (+)-side cap is created, with a (blue) double point loop.(d) After tubing two (−)-side cap intersections into the transverse grope, a new intersection between (−)-side caps is created, with a blue double point loop.

the double point loops are also mapped to $1_{\Gamma'}$ under ϕ. Thus, Γ' is good.

For (2), let Γ be a good group and let $\Gamma' \lhd \Gamma$ be a normal subgroup. Suppose that $\phi\colon \pi_1(G) \to \Gamma/\Gamma'$ is a homomorphism. Define a homomorphism $\widetilde{\phi}\colon \pi_1(G) \to \Gamma$ by setting $\widetilde{\phi}(a_i) = x_i$ for some choice of x_i with $\phi(a_i) = x_i\Gamma'$. By definition, $p \circ \widetilde{\phi} = \phi$ where $p\colon \Gamma \to \Gamma/\Gamma'$ is the quotient map. Since Γ is good, we obtain a disc $D \looparrowright G$ such that the double point loops are mapped to 1_Γ under $\widetilde{\phi}$. Thus, the double point loops are mapped to $1_{\Gamma/\Gamma'}$ by $\phi = p \circ \widetilde{\phi}$ which implies that Γ/Γ' is good.

For (3), let

$$1 \longrightarrow \Gamma_1 \xrightarrow{\ f\ } \Gamma \xrightarrow{\ g\ } \Gamma_2 \longrightarrow 1$$

be an exact sequence of groups, where Γ_1 and Γ_2 are good. Let $\phi\colon \pi_1(G) \to \Gamma$ be a homomorphism. Perform grope height raising (Proposition 17.3) and then apply the sequential contraction lemma (Lemma 17.7) to obtain a height 4 capped grope G' with the same attaching region as G and whose caps are mutually disjoint.

Removing the bottom height 2 grope from G' leaves us with a pairwise disjoint collection of height 2 disc-like capped gropes. By contracting, produce a mutually disjoint collection $\{G_i''\}$ of height 1.5 disc-like capped gropes. For each i, we get a map $\pi_1(G_i'') \xrightarrow{\iota_*} \pi_1(G) \xrightarrow{\phi} \Gamma \xrightarrow{g} \Gamma_2$, where ι is the inclusion $G_i'' \hookrightarrow G$.

Since Γ_2 is good, there exist immersed discs $D_i \looparrowright G_i''$, whose framed boundary coincides with the attaching region of G_i'' such that the double point loops are mapped by $g \circ \phi \circ \iota_*$ to 1_{Γ_2}. Attach these discs to the bottom height 2 grope of G' to obtain a height 2 disc-like capped grope. Contract further to produce a height 1.5 disc-like capped grope, which we call $\overline{G}$. Note that $\overline{G}$ has the same (framed) attaching region as the original capped grope G. Since the collection $\{G_i''\}$ is mutually disjoint, the double point loops of the caps of $\overline{G}$ generate $\pi_1(\overline{G})$, and thus the map

$$\pi_1(\overline{G}) \xrightarrow{\iota_*} \pi_1(G) \xrightarrow{\phi} \Gamma \xrightarrow{g} \Gamma_2$$

is trivial, where ι now denotes the inclusion $\overline{G} \hookrightarrow G$. Thus the image of $\pi_1(\overline{G}) \xrightarrow{\iota_*} \pi_1(G) \xrightarrow{\phi} \Gamma$ lies in the kernel of g, which by hypothesis lies in the image of the injective map f. It follows that we have a map $\pi_1(\overline{G}) \xrightarrow{\iota_*} \pi_1(G) \xrightarrow{\widetilde{\phi}} \Gamma_1$ where $\widetilde{\phi} \circ f = \phi$.

Since Γ_1 is good, there is an immersed disc $D \looparrowright \overline{G}$ such that the double point loops are mapped by $\widetilde{\phi} \circ \iota_*$ to 1_{Γ_1}. Since $\widetilde{\phi} \circ f = \phi$, we have found D with framed boundary coinciding with the attaching region of G such that the double point loops are mapped by ϕ to 1_Γ as needed. Thus Γ is good.

For (4), let I be a directed set and let Γ be the colimit of the directed system of groups $\{\Gamma_\alpha\}_{\alpha \in I}$ with the corresponding family of homomorphisms $\{\psi_{\alpha\beta}\colon \Gamma_\alpha \to \Gamma_\beta\}_{\alpha,\beta \in I}$, where each Γ_α is good. By definition, there exist maps $\{\psi_\alpha\colon \Gamma_\alpha \to \Gamma\}_{\alpha \in I}$ such that $\psi_\alpha = \psi_\beta \circ \psi_{\alpha\beta}$ for all $\alpha, \beta \in I$. Let $\phi\colon \pi_1(G) \to \Gamma$ be a homomorphism. Since Γ is a colimit, for each i there is an $\alpha_i \in I$ such that the double point loop a_i satisfies $\phi(a_i) = \psi_{\alpha_i}(\overline{a}_i)$ for some $\overline{a}_i \in \Gamma_{\alpha_i}$. Let γ be such that $\alpha_i \leq \gamma$ for all i. Then $\phi(a_i) = \psi_\gamma(\widetilde{a}_i)$ for some $\widetilde{a}_i \in \Gamma_\gamma$. Define $\widetilde{\phi}\colon \pi_1(G) \to \Gamma_\gamma$ by mapping $a_i \mapsto \widetilde{a}_i$. By definition, $\psi_\gamma \circ \widetilde{\phi} = \phi$. Then, since Γ_γ is good, there exists an immersed disc $D \looparrowright G$ with the desired framed boundary whose double point loops are mapped by $\widetilde{\phi}$ to 1_{Γ_γ}. Since $\psi_\gamma \circ \widetilde{\phi} = \phi$, the double point loops of D are mapped by ϕ to 1_Γ. Thus Γ is good.

For (5), let Γ be a group and let $\Gamma' \leq \Gamma$ such that Γ' is good and $[\Gamma : \Gamma'] < \infty$. Let Γ'' be the normal core of Γ', that is $\Gamma'' \lhd \Gamma$, $\Gamma'' \leq \Gamma'$, and $[\Gamma : \Gamma''] < \infty$. Since $\Gamma'' \leq \Gamma'$ and Γ' is good, we see that Γ'' is good by (1). Moreover, we have the short exact sequence

$$1 \longrightarrow \Gamma'' \longrightarrow \Gamma \longrightarrow \Gamma/\Gamma'' \longrightarrow 1$$

where Γ/Γ'' is a finite group by hypothesis. By (3) and the fact that finite groups are good (Theorem 19.2), the group Γ is good. □

Example 19.6. The combination of our work so far produces several classes of good groups as follows. We immediately see that any finitely generated abelian group is good by Theorems 19.2 and 19.4 together with Operation (3). Moreover, since every group is a colimit of its finitely generated subgroups, every abelian group is good by Operation (4). Solvable groups are good by Operation (3) since they are constructed, by definition, as the result of a finite sequence of extensions by abelian groups. Nilpotent groups are good since they are all solvable. More generally, the members of the class generated by finite groups and abelian groups, along with Operations (1)–(4), namely the *elementary amenable groups*, are good. Recall that the adjective *virtual* applied to a property of a group means that the property holds for a finite index subgroup. Operation (5) implies, for example, that virtually solvable and virtually polycyclic groups are also good.

We also have the following result, showing that the goodness of the 2-generator free group $\mathbb{Z} * \mathbb{Z}$ is of particular interest. Whether or not free groups are good remains open, as we discuss in detail in Chapter 23.

Proposition 19.7. *All groups are good if and only if $\mathbb{Z} * \mathbb{Z}$ is good.*

Proof. One direction is trivial. For the other, note that every finitely generated group arises as a quotient of a subgroup of $\mathbb{Z} * \mathbb{Z}$ and that any group is a colimit of its finitely generated subgroups. By Operations (1), (2), and (4), every group is good if $\mathbb{Z} * \mathbb{Z}$ is good. □

Since Freedman's original work, the idea of proving that a group is good using its growth rate, specifically by understanding the effect of grope height raising and the contraction and push off operations on double point loops, has been extended. To describe these results, we will need the following definition.

Definition 19.8. For S a finite subset of a group Γ, the *growth function* $g_S \colon \mathbb{N} \to \mathbb{N}$ maps the integer r to the number of distinct elements in Γ that can be written as words of length $\leq r$ in the elements of S and their inverses.

A group Γ has *polynomial growth* if for all finite subsets $S \subseteq \Gamma$, there is a polynomial f such that $g_S(r) \leq f(r)$ for sufficiently large r. A group Γ has *subexponential growth* if $g_S(r) \leq b^r$ for sufficiently large r for any finite subset $S \subseteq \Gamma$ and real number $b > 1$. A group Γ has *exponential growth* if it does not have subexponential growth.

We have seen in Example 19.6 that virtually nilpotent groups are good. Gromov proved that a finitely generated group has polynomial growth if and only if it is virtually nilpotent [**Gro81**]. It follows that any finitely generated group of polynomial growth is good. The following theorem is a significant improvement of this fact. Together with Proposition 19.5, it provides the largest known class of good groups.

Theorem 19.9 ([**FT95a**, **KQ00**]). *Every group of subexponential growth is good.*

The proof by Freedman and Teichner [**FT95a**] (see also the erratum as an appendix in [**KQ00**]) is an improvement on the techniques given in [**FQ90**, Section 2.9] and recounted in Theorems 19.2 and 19.4 above, consisting of performing 'linear' grope height raising, followed by 'exponential' contraction and subsequent push off. The proof by Krushkal and Quinn [**KQ00**] uses a 'grope-splitting' technique. We shall not describe these proofs further.

At the time of writing, every known finitely presented subexponential growth group is elementary amenable, so Theorem 19.9 is not known to be an improvement for compact manifolds. Nonetheless, the above theorem produces many examples of good groups that are not elementary amenable, as follows. Grigorchuk proved that there are uncountably many groups of subexponential growth which are not of polynomial growth [**Gri85**]. These groups are good by Theorem 19.9, but are not elementary amenable since it was shown by Chou [**Cho80**] (generalising the Milnor-Wolf theorem [**Mil68**, **Wol68**] for solvable groups) that elementary amenable groups have either polynomial or exponential growth.

Having subexponential growth is not a necessary condition for goodness. Consider the Baumslag-Solitar groups

$$BS(1,n) = \langle a, b \mid bab^{-1} = a^n \rangle.$$

Note that $BS(1,n)$ is solvable, and hence good, for all n. However, we show now that $BS(1,n)$ has exponential growth when $n \geq 2$. Milnor [**Mil68**] showed that every solvable group which is not polycyclic has exponential growth. When $n \geq 2$, $BS(1,n)$ is not polycyclic, since its commutator subgroup is the infinitely generated abelian group $\mathbb{Z}[\frac{1}{n}]$ and any subgroup of a polycyclic group must be finitely generated. Thus, $BS(1,n)$ has exponential growth, when $n \geq 2$.

Groups of subexponential growth are in particular amenable (see, for example, [**Gre69**]). The class of amenable groups is also closed under the operations (1)–(5) of Proposition 19.5, so all solvable groups are amenable. Such similarities between the classes of amenable groups and good groups suggest that the extension of Theorem 19.9 to amenable groups is plausible, and this is currently an interesting open problem.

CHAPTER 20

The s-cobordism theorem, the sphere embedding theorem, and the Poincaré conjecture

Patrick Orson, Mark Powell, and Arunima Ray

We start this chapter by finishing the proof of the s-cobordism theorem begun in Chapter 1. This will allow us to prove the category losing version of the 4-dimensional Poincaré conjecture (Theorem 20.3), that every smooth homotopy 4-sphere is homeomorphic to S^4.

Then we will prove an alternative statement of the disc embedding theorem [**FQ90**, Theorem 5.1B]. This version differs from the usual disc embedding theorem in that the intersection conditions are on the immersed discs $\{f_i\}$ rather than on the dual spheres $\{g_i\}$, and that the embedded discs $\{\overline{f}_i\}$ obtained as a consequence of the theorem are regularly homotopic to the original immersed discs $\{f_i\}$. We show that the two versions are logically equivalent. We directly obtain the *sphere embedding theorem* from the second version. The sphere embedding theorem is the key ingredient needed to deduce the exactness of the surgery sequence, as we show in Chapter 22. For the applications to surgery, it will be important that we have geometrically transverse spheres in the outcome of the sphere embedding theorem.

20.1. The s-cobordism theorem

In this section, we state and prove the s-cobordism theorem for smooth s-cobordisms between closed 4-manifolds. As indicated in Chapter 1, the disc embedding theorem is a key ingredient in the proof.

Theorem 20.1 (s-cobordism theorem). *Let N be a smooth 5-dimensional h-cobordism between closed 4-manifolds M_0 and M_1 with vanishing Whitehead torsion $\tau(N, M_0)$. Further, suppose that $\pi_1(N)$ is a good group. Then N is homeomorphic to the product $M_0 \times [0,1]$. In particular, M_0 and M_1 are homeomorphic.*

Proof. The first part of the proof was already explained in Chapter 1. We give the argument again, with a few more details, for the convenience of the reader. First, since N is smooth, we may choose a Morse function $F\colon N \to [0,1]$ with $F^{-1}(i) = M_i$ for $i = 0, 1$. This gives rise to a handle decomposition of N relative to M_0, as defined in Chapter 13. As usual, by transversality, we may assume that the handles are attached in increasing order of index. Since N is connected, we may assume, by handle cancellation, that this handle decomposition has no 0- or 5-handles.

Next we use the process of *handle trading* to trade 1- and 4-handles for 3- and 2-handles respectively, as follows. Let $N_2 \subseteq N$ denote the union of $M_0 \times [0,1]$ and the 1- and 2-handles of N. Let M_2 denote the new boundary, so $\partial N_2 = -M_0 \sqcup M_2$.

Consider the chain of inclusion induced maps $\pi_1(M_0) \to \pi_1(N_2) \to \pi_1(N)$. Since N is built from N_2 by attaching handles of index strictly greater than 2, the second map is an isomorphism. The composition is an isomorphism by hypothesis. Thus the first map is an isomorphism.

Fix a 1-handle h^1 in N_2, with core arc α. We claim that there is an arc $\beta \subseteq M_0$ such that $\gamma := \alpha \cup \beta$ is a null-homotopic loop in N_2. To see this, first choose any arc β' with the same endpoints as α. Then there is some loop $\delta \subseteq M_0$ with the same image in $\pi_1(N_2)$ as $\alpha \cup \beta'$, since the inclusion induced map $\pi_1(M_0) \to \pi_2(N_2)$ is surjective. The connected sum of β' and δ^{-1} is the desired β. By transversality, we assume that γ is disjoint from the attaching circles of all the 1- and 2-handles of N_2 and then we push γ to the boundary M_2.

By turning handles upside down, we see that the inclusion induced map $\pi_1(M_2) \to \pi_1(N_2)$ is an isomorphism. Thus γ bounds an immersed disc in M_2, since it is null-homotopic in N_2. By finger moves in the direction of γ, we see that γ bounds an embedded disc in M_2. Thicken this disc to produce a cancelling 2-/3-handle pair. More precisely, insert a collar of $M_2 \times [0,1]$ into the handle decomposition and thicken by pushing the interior of the disc into this collar. The result is the addition of a single cancelling 2-/3-handle pair compatible with the old handle decomposition. By the choice of γ the 2-handle cancels the 1-handle h^1, leaving the 3-handle behind. Iterating this process allows us to trade all the 1-handles in N for 3-handles, and the same argument for the dual handlebody of N built on M_1 trades all 4-handles for 2-handles.

At this point, we have produced a handle decomposition of N, relative to M_0, consisting only of 2- and 3-handles, attached in that order. Let $N_{1/2}$ denote the 5-manifold consisting of M_0 and the 2-handles, and let $M_{1/2}$ denote the 4-manifold obtained as a result. That is, $\partial N_{1/2} = -M_0 \sqcup M_{1/2}$. Then the inclusion induced map $\pi_1(M_{1/2}) \to \pi_1(N_{1/2})$ is an isomorphism since $N_{1/2}$ is produced from $M_{1/2}$ by attaching only 3-handles. We also know that the inclusion-induced map $\pi_1(N_{1/2}) \to \pi_1(N)$ is an isomorphism since N is produced from $N_{1/2}$ by adding only 3-handles. Thus, the inclusion induced map $\pi_1(M_{1/2}) \to \pi_1(N)$ is an isomorphism and can be used to identify the two groups.

We obtain a chain complex corresponding to the handle decomposition of N constructed above of the form

$$0 \to C_3(\widetilde{N}, \widetilde{M_0}) \xrightarrow{\partial_3} C_2(\widetilde{N}, \widetilde{M_0}) \to 0,$$

where $\widetilde{N}$ is the universal cover of N and $\widetilde{M_0}$ is the induced cover of M_0, which in this case is the universal cover since $\pi_1(M_0) \to \pi_1(N)$ is an isomorphism. Each $C_i(\widetilde{N}, \widetilde{M_0})$ is a finitely generated, free $\mathbb{Z}[\pi_1(N)]$-module with basis elements corresponding to a choice of lift of the i-handles of N. The boundary map records the intersections, with $\mathbb{Z}[\pi_1(N)] = \mathbb{Z}[\pi_1(M_{1/2})]$ coefficients, between the belt spheres $\{S_1, \dots, S_k\}$ of the 2-handles, corresponding to $\{0\} \times S^2 \subseteq D^2 \times D^3$, and the attaching spheres $\{T_1, \dots, T_k\}$ of the 3-handles, corresponding to $S^2 \times \{0\} \subseteq D^3 \times D^2$. Each of the sets $\{S_i\}$ and $\{T_i\}$ is a collection of pairwise disjoint, framed, embedded 2-spheres in $M_{1/2}$.

The vanishing of the Whitehead torsion $\tau(N, M_0)$ implies that after possible stabilisation by adding cancelling pairs of 2- and 3-handles, and handle slides (corresponding to basis changes), the boundary map ∂_3 is represented by the identity matrix [**Lüc02**, Chapter 2]. In other words, we may assume that $\lambda(S_i, T_j) = \delta_{ij}$, measured in $\mathbb{Z}[\pi_1(M_{1/2})]$, for all i, j. As in the introduction, we wish to perform an isotopy of the family $\{T_i\}$ such that these intersection numbers are realised geometrically, so that we can cancel the 2-handles with the 3-handles.

Since the inclusion induced map from $\pi_1(M_0)$ to $\pi_1(N)$ is an isomorphism, the 2-handles in N are attached to M_0 along homotopically trivial circles in M_0. These null homotopies, glued to the cores of the attached 2-handles (pushed to the boundary), produce a collection $\{S_i^{\#}\}$ of possibly unframed, immersed spheres in $M_{1/2}$, such that the collections $\{S_i\}$ and $\{S_i^{\#}\}$ are geometrically transverse. The identical argument applied to the dual handlebody obtained by turning the handles upside down, produces the family $\{T_i^{\#}\}$ of possibly unframed, immersed spheres in $M_{1/2}$ geometrically transverse to $\{T_i\}$.

Note that we do not have any control over the intersections between the families $\{S_i^{\#}\}$ and $\{T_i\}$, nor between the families $\{S_i\}$ and $\{T_i^{\#}\}$. There are also uncontrolled intersections within and between the families $\{S_i^{\#}\}$ and $\{T_i^{\#}\}$.

We will now arrange for framed, geometrically transverse spheres for $\{S_i\}$, and respectively for $\{T_i\}$, that are disjoint from $\{T_i\}$, respectively $\{S_i\}$. This will deviate slightly from the proof sketched in Chapter 1, since we desire framed dual spheres rather than the unframed ones we produced there.

Let $\{S_i'\}$ and $\{T_i'\}$ denote parallel copies of $\{S_i\}$ and $\{T_i\}$, respectively. Then, by hypothesis, the collection $\{S_i\} \cup \{S_i'\}$ is pairwise disjoint, as is the collection $\{T_i\} \cup \{T_i'\}$ (recall that $\{S_i\}$ and $\{T_i\}$ are collections of mutually disjoint, embedded spheres). Additionally, $\lambda(S_i', T_j) = \lambda(S_i, T_j') = \delta_{ij}$ for all i, j. Thus all but one intersection point between S_i' and T_i, as well as between S_i and T_i', can be paired by immersed Whitney discs, for each i. In addition, all the intersection points between S_i' and T_j, as well as between S_i and T_j', can be paired by immersed Whitney discs, for all $i \neq j$. Let $\{W_\ell^S\}$ and $\{W_\ell^T\}$ be the collections of these Whitney discs for the extraneous intersections between $\{S_i\}$ and $\{T_j'\}$, and between $\{S_i'\}$ and $\{T_j\}$, respectively.

We make the interiors of the $\{W_\ell^T\}$ disjoint from $\{T_i\}$ by tubing into the unframed geometrically transverse spheres $\{T_i^{\#}\}$. This creates new intersections of the Whitney discs with $\{S_i\}$, but not with $\{T_i\}$, since $\{T_i^{\#}\}$ and $\{T_i\}$ are geometrically transverse. Then remove all intersections of the interiors with $\{S_i\}$ by (disjoint) finger moves in the direction of $\{T_i\}$. See Figure 20.1. Boundary twist, at the expense of new intersections with $\{S_i'\}$, to correct the framing of the Whitney discs. We still call the resulting collection of Whitney discs $\{W_\ell^T\}$, but note that they are framed and their interiors lie in the complement of $\bigcup\{S_i\} \cup \bigcup\{T_i\}$. We do not control the intersections within the collection $\{W_\ell^T\}$. We have created some new algebraically cancelling intersections between $\{S_i\}$ and $\{T_i\}$ via the finger move. Instead of finger moving the $\{S_i\}$ over $\{T_i\}$, we could have finger moved T_i over S_i, with isotopic results. So we shift perspective and consider these finger moves to be an isotopy of $\{T_i\}$ instead.

A similar process makes the interiors of the $\{W_\ell^S\}$ disjoint from $\bigcup\{S_i\} \cup \bigcup\{T_i\}$. That is, remove intersections of the discs with $\{S_i\}$ by tubing into the unframed transverse spheres $\{S_i^{\#}\}$, next remove all intersections with $\{T_i\}$ by disjoint finger moves in the direction of $\{S_i\}$, and then boundary twist at the expense of new intersections with $\{T_i'\}$, to frame the Whitney discs. We still call the resulting collection of framed, immersed Whitney discs $\{W_\ell^S\}$.

Perform the Whitney move on $\{S_i'\}$ along the Whitney discs $\{W_\ell^T\}$, and call the resulting spheres $\{\widehat{T}_i\}$. Note that the collections $\{T_i\}$ and $\{\widehat{T}_i\}$ are geometrically transverse, and moreover, $S_i \cap \widehat{T}_j = \emptyset$ for all i, j. Then perform the Whitney move on $\{T_i'\}$ along the Whitney discs $\{W_\ell^S\}$ and call the resulting spheres $\{\widehat{S}_i\}$. As desired, the collections $\{S_i\}$ and $\{\widehat{S}_i\}$ are geometrically transverse, and $\widehat{S}_i \cap T_j = \emptyset$ for all i, j. Thus the collections $\bigcup\{S_i\} \cup \bigcup\{T_i\}$ and $\bigcup\{\widehat{S}_i\} \cup \bigcup\{\widehat{T}_i\}$ are geometrically transverse, and as a result the collection $\bigcup\{S_i\} \cup \bigcup\{T_i\}$ is π_1-negligible in $M_{1/2}$.

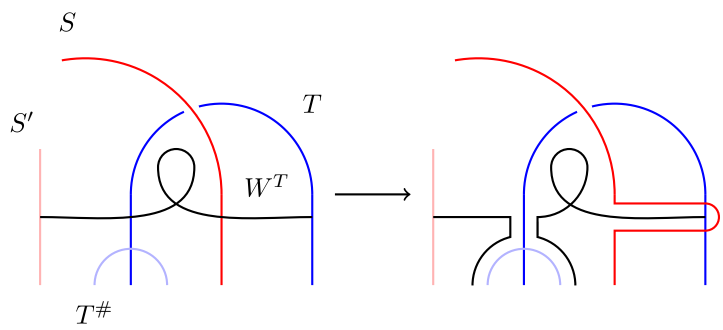


Figure 20.1. Obtaining a Whitney disc for intersections between $\{S_i'\}$ and $\{T_i\}$ with interior in the complement of $\bigcup\{S_i\} \cup \bigcup\{T_i\}$. We see a Whitney disc W^T (black) pairing intersection points between S' (light red) and T (blue). Remove intersections of W^T and T by tubing into the unframed geometric dual $T^{\#}$ (light blue). Intersections of W^T and S, including any new ones created in the previous step, can be removed by a finger move in the direction of T.

We also remark that in the process of constructing $\bigcup\{\widehat{S}_i\} \cup \bigcup\{\widehat{T}_i\}$, we have moved the $\{T_i\}$ by an isotopy, the collection $\{S_i\}$ is unaffected, and we have created some new algebraically cancelling intersection points between the collections $\{S_i\}$ and $\{T_i\}$.

Now we return to our original problem, which is to perform a further isotopy of the $\{T_i\}$ so that the collections $\{S_i\}$ and $\{T_i\}$ become geometrically transverse. We have that $\lambda(S_i, T_j) = \delta_{ij}$ for all i, j. Thus all the unwanted double points between $\{S_i\}$ and $\{T_i\}$ can be paired up by framed, immersed Whitney discs $\{W_m\}$ in $M_{1/2}$. For each intersection of the interiors of $\{W_m\}$ with $\{S_i\}$, tube $\{W_m\}$ into the geometrically transverse spheres $\{\widehat{S}_i\}$. Similarly, for every intersection of the interiors of $\{W_m\}$ with $\{T_i\}$, tube $\{W_m\}$ into the geometrically transverse spheres $\{\widehat{T}_i\}$. Since $\{\widehat{S}_i\}$ and $\{\widehat{T}_i\}$ are framed, the collection $\{W_m\}$ remains framed and we have now arranged that the interior of $\{W_m\}$ lies in the complement of $\bigcup\{S_i\} \cup \bigcup\{T_i\}$.

Our goal is to apply the disc embedding theorem to the collection $\{W_m\}$ in this complement. With this in mind, we need to construct algebraically transverse spheres for $\{W_m\}$.

The desired spheres will arise from Clifford tori. Let Σ_m be the Clifford torus at one of the two double points paired by some W_m. For each m, the Clifford torus Σ_m intersects W_m exactly once, and the collection of such Clifford tori is embedded and pairwise disjoint. Cap each Σ_m with the meridional discs to $\{S_i\}$ and $\{T_i\}$ described in Figure 20.2. Each cap has a unique intersection with $\bigcup\{S_i\} \cup \bigcup\{T_i\}$, and none of them intersects $\{W_m\}$. Tube these intersections into parallel copies of the relevant members of the set of geometrically transverse spheres $\{\widehat{S}_i\} \cup \{\widehat{T}_i\}$. Contract these capped surfaces to produce algebraically transverse spheres $\{R_m\}$ for the discs $\{W_m\}$. Since the collection $\{R_m\}$ is produced by contraction of capped surfaces with mutually disjoint bodies, each element is framed and we see that $\lambda(R_m, R_{m'}) = 0 = \mu(R_m)$ for all m, m'.

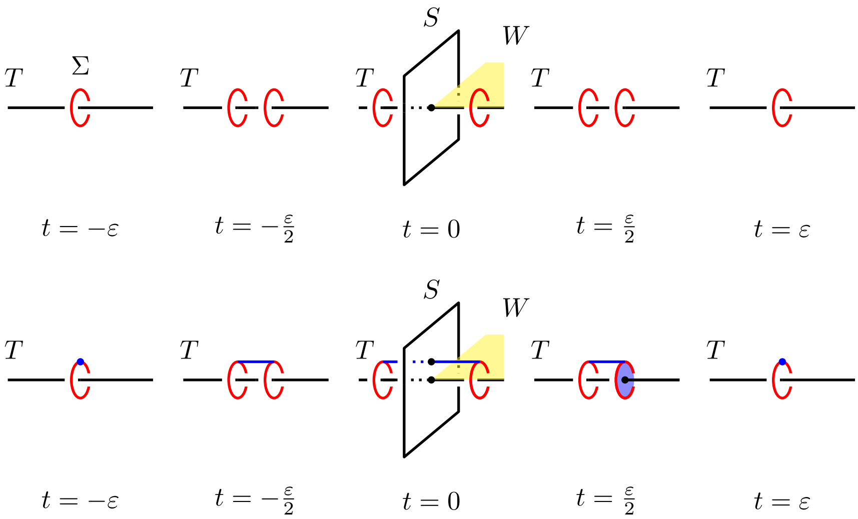


Figure 20.2. Obtaining a transverse sphere from a Clifford torus. Top: The Clifford torus Σ (red) at one of the two intersection points paired up by the Whitney disc W (yellow). The single point of intersection between T and W is shown in the central panel. Bottom: The two meridional discs are shown in blue. We see that each meridional disc intersects exactly one of $\{S, T\}$, exactly once.

The Whitney discs $\{W_m\}$, along with the collection of spheres $\{R_m\}$, now satisfy the hypotheses of the disc embedding theorem in the 4-manifold

$$M' := M_{1/2} \smallsetminus \left(\bigcup \nu S_i \cup \bigcup \nu T_i \right).$$

Since $\{S_i\} \cup \{T_i\}$ is π_1-negligible in $M_{1/2}$, we see that $\pi_1(M') \cong \pi_1(M_{1/2}) \cong \pi_1(N)$, which is a good group by hypothesis. Additionally, the intersection numbers of $\{R_m\}$ vanish in M' since they vanish in $M_{1/2} \supseteq M'$ and the inclusion map induces a π_1-isomorphism.

The disc embedding theorem replaces the Whitney discs $\{W_m\}$ by embedded discs with the same framed boundaries. (We also obtain geometrically transverse spheres for these embedded discs, but we will not need them here.) Perform Whitney moves on the $\{T_i\}$ using the framed, embedded Whitney discs to remove all the unwanted intersections. This is the desired isotopy of $\{T_i\}$, after which the collections $\{S_i\}$ and $\{T_i\}$ become geometrically transverse. Now the 2-handles and the 3-handles of the 5-manifold N can be cancelled in pairs. Since there are no remaining handles, N is homeomorphic to the product $M_0 \times [0, 1]$, as desired. □

The h-cobordism theorem is an immediate corollary of the s-cobordism theorem, since the Whitehead torsion of a simply connected cobordism lies in the Whitehead group of the trivial group, which is trivial.

Theorem 20.2 (h-cobordism theorem). *Every smooth h-cobordism between simply connected, closed 4-manifolds M_0 and M_1 is homeomorphic to the product $M_0 \times [0, 1]$.*

20.2. The Poincaré conjecture

Possibly the most famous application of the disc embedding theorem is the 4-dimensional Poincaré conjecture.

THEOREM 20.3 (Poincaré conjecture, category losing version). *Every closed, smooth 4-manifold homotopy equivalent to the 4-sphere S^4 is homeomorphic to S^4.*

The proof we give below uses 5-dimensional surgery and the h-cobordism theorem (Theorem 20.2). We discuss the category preserving Poincaré conjecture, that every topological homotopy 4-sphere is homeomorphic to S^4, in Section 21.6.2. As explained there, the known proofs require ingredients not proved in this book, such as the category preserving h-cobordism theorem.

PROOF. Let Σ be a closed, smooth 4-manifold homotopy equivalent to the 4-sphere S^4. The signature of Σ vanishes since $H_2(\Sigma;\mathbb{Z}) = 0$. We claim that the tangent bundle of Σ is stably trivial. The obstructions to stably trivialising the tangent bundle of a smooth, oriented 4-manifold are the second Stiefel-Whitney class $w_2(T\Sigma)$ and the first Pontryagin class $p_1(T\Sigma)$. Since the cohomology of Σ is concentrated in degree four, $w_2(T\Sigma) = 0$, while $p_1(T\Sigma)$ vanishes because the signature vanishes, by the Hirzebruch signature formula $3\sigma(\Sigma) = \langle p_1(T\Sigma), [\Sigma]\rangle \in \mathbb{Z}$. Thus the tangent bundle is stably trivial as claimed. It follows that Σ bounds a compact 5-manifold W with stably trivial tangent bundle, since the smooth framed 4-dimensional cobordism group Ω_4^{fr} is trivial.

Construct a degree one normal map relative boundary $(f,\partial f)\colon (W,\Sigma) \to (D^5, S^4)$ by collapsing to a point the complement of an open tubular neighbourhood in W of some point $x \in \Sigma$. More precisely, the map ∂f is a homotopy equivalence and f sends fundamental class to fundamental class. We use the fact that W is stably framed to construct the normal data required for the normal map. For more details on normal maps, see Section 22.1.4.

Perform 5-dimensional surgery on $(f,\partial f)$ to make f into a homotopy equivalence. This is possible since the odd dimensional surgery obstruction group $L_5(\mathbb{Z})$ of the trivial group is itself the trivial group. Thus Σ bounds a smooth, contractible 5-manifold W'. In this step we have used the main result of odd-dimensional surgery theory [**Wal99**, Theorem 6.4]; see also Section 22.1.6 below for the definition of the group $L_5(\mathbb{Z}) \cong L_5^s(\mathbb{Z})$. The proof so far shows that any smooth homology 4-sphere bounds a smooth, contractible 5-manifold.

Remove an open ball from the interior of W'. This produces an h-cobordism from Σ to S^4. By the h-cobordism theorem, W' is homeomorphic to the product $S^4 \times [0,1]$. Consequently Σ is homeomorphic to S^4. □

In the previous proof, the h-cobordism between Σ and S^4 could have been obtained using Wall's theorem [**Wal64**] showing that any two closed, smooth, simply connected 4-manifolds with isomorphic intersection forms are smoothly h-cobordant. The proof given here is more transparent and has the advantage that it applies, with a few modifications, to topological homotopy 4-spheres, as we describe in Section 21.6.2.

20.3. The sphere embedding theorem

We state and prove an alternative version of the disc embedding theorem, where the intersection assumptions are on the initial immersed discs rather than the dual spheres.

THEOREM 20.4 ([**FQ90**, Theorem 5.1B]). *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 boundaries satisfying $\mu(f_i) = 0$ for all i and $\lambda(f_i,f_j) = 0$ for all $i \neq j$. Suppose moreover that there is *an immersed collection*

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

of framed dual spheres, that is $\lambda(f_i, g_j) = \delta_{ij}$ *for all* $i, j = 1, \dots, k$.

Then there exist mutually disjoint flat embeddings

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

with $\overline{f}_i$ *regularly homotopic relative boundary to* f_i *for each* i*, together with an immersed collection of framed geometrically transverse spheres*

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

such that, for each i, $\overline{g}_i$ *is homotopic to* g_i.

PROOF. Since $\lambda(f_i, g_j) = \delta_{ij}$, we may apply the geometric Casson lemma (Lemma 15.3) to arrange that $\{f_i\}$ and $\{g_i\}$ are geometrically transverse (note that the collection of transverse spheres $\{g_i\}$ may have any kind of intersections among themselves). This changes the collections by regular homotopy, and we continue to use the same notation. Since $\lambda(f_i, f_j) = 0$ for all $i \neq j$ and $\mu(f_i) = 0$ for all i, the intersections and self-intersections within $\{f_i\}$ are paired by framed, immersed Whitney discs.

Consider one such Whitney disc D pairing up intersections between f_i and f_j, where possibly $i = j$. Such a disc may intersect itself, the collections $\{f_i\}$ and $\{g_i\}$, or other Whitney discs (see the left panel of Figure 20.3). For each intersection of D with f_ℓ, for some ℓ, tube D into a parallel push-off of the geometric dual g_ℓ, as shown in the right panel of Figure 20.3. This introduces potentially many new in-

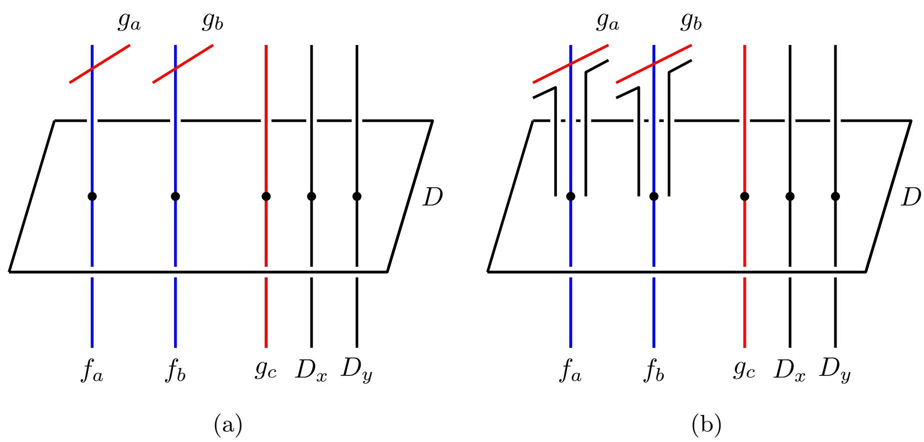


Figure 20.3. Left: A schematic picture of a piece of a Whitney disc D. It may intersect $\{f_i\}$, $\{g_i\}$, or other Whitney discs. Recall that $\{f_i\}$ and $\{g_i\}$ are geometrically transverse. Right: Remove intersections of D with $\{f_i\}$ by tubing into $\{g_i\}$.

tersections, between D and anything that intersected g_ℓ (including g_ℓ itself), as well as new self-intersections of D coming from the self-intersections of g_ℓ. However, the interior of D no longer intersects any f_i, since g_ℓ intersects exactly one of the $\{f_i\}$, namely f_ℓ, at the intersection point we used for tubing. Do this for all the Whitney discs and their intersections with $\{f_i\}$. Now our Whitney discs are more complicated, but their interiors lie in the complement of $\bigcup\{f_i\}$. Call this collection of Whitney discs $\{D'_k\}$. These Whitney discs are framed, so if they were embedded we could perform the Whitney move along them to obtain the embedded discs we seek.

We wish to apply the disc embedding theorem to $N := M \smallsetminus \bigcup \nu f_i$. Since each f_i has a geometrically transverse sphere by construction, the collection $\{f_i\}$ is π_1-negligible, and so there is an isomorphism $\pi_1(N) \to \pi_1(M)$. Since $\pi_1(M)$ is good, we conclude that $\pi_1(N)$ is also good, as desired.

Next, we find algebraically transverse spheres for the Whitney discs $\{D'_k\}$. As before, these will arise from Clifford tori. Let Σ_k be the Clifford torus at one of the two double points paired by some D'_k. As we saw earlier, the Clifford torus Σ_k intersects D'_k exactly once, and the collection of such Clifford tori are framed, embedded and pairwise disjoint. Cap each Σ_k with meridional discs to $\{f_i\}$ (see Figure 20.2). Each cap has a unique intersection with $\{f_i\}$, and none intersects $\{D'_k\}$. Tube these intersections between the caps and $\{f_i\}$ into the set of geometrically transverse spheres $\{g_i\}$ and contract the resulting capped surfaces in the complement of $\{f_i\}$ to produce algebraically transverse spheres $\{g'_k\}$ for the discs $\{D'_k\}$ lying in N. Since the collection $\{g'_k\}$ is produced by contraction of capped surfaces with disjoint bodies, each g'_k is framed and $\lambda(g'_k, g'_\ell) = 0 = \mu(g'_k)$ in N for all k, ℓ. Moreover note that by Lemma 17.11, we have $[g'_k] = 0 \in \pi_2(M)$ for each k.

We may therefore apply the disc embedding theorem to replace the immersed Whitney discs $\{D'_k\}$ with topologically embedded Whitney discs $\{W_k\}$ with normal bundles that induce the right framing on the boundaries, and framed, geometrically transverse spheres $\{R_k\}$ in N, with R_k homotopic to g'_k for each j. For each intersection of some g_i with some W_k, tube that g_i into the geometrically transverse sphere R_k. This transforms the collection $\{g_i\}$ to a collection $\{\overline{g}_i\}$, which may have more intersections among themselves, but are still geometrically transverse to $\{f_i\}$. The $\{\overline{g}_i\}$ are still framed because the $\{R_k\}$ are. Since g'_k is null-homotopic in M for every j, so is R_k. It follows that $\overline{g}_i$ is homotopic to g_i for each i.

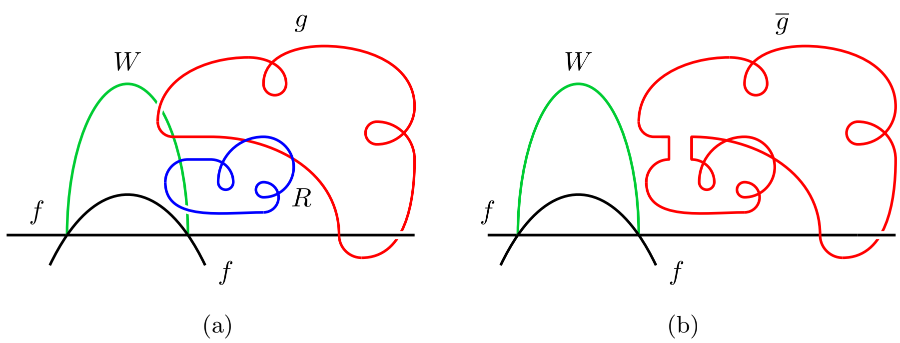


Figure 20.4. Left: An embedded Whitney disc (green) with a geometrically dual sphere (blue), both with interiors in the complement of $\{f_i\}$, has been produced.
Right: After tubing $\{g_i\}$ into the geometrically dual spheres as needed we have produced the spheres $\{\overline{g}_i\}$, which are geometrically dual to $\{f_i\}$.

Moreover, we obtain embedded, flat, framed Whitney discs for the intersections among the $\{f_i\}$ in $M \smallsetminus (\bigcup \nu f_i \cup \bigcup \nu \overline{g}_i)$ (see Figure 20.4). Perform the Whitney move on $\{f_i\}$ over the Whitney discs $\{W_j\}$ to obtain flat, embedded discs $\{\overline{f}_i\}$, regularly homotopic to the $\{f_i\}$, with the same framed boundary as the $\{f_i\}$, as well as framed, geometrically transverse spheres $\{\overline{g}_i\}$. □

PROPOSITION 20.5. *Theorem 20.4 and the disc embedding theorem are equivalent.*

PROOF. Since we already deduced Theorem 20.4 from the disc embedding theorem, it suffices to show the converse. Begin with immersed discs $\{f_i\}$ with algebraically transverse spheres $\{g_i\}$ with $\lambda(f_i, g_j) = \delta_{ij}$ and $\lambda(g_i, g_j) = \mu(g_i) = 0$ for all i, j. Tube each intersection and self-intersection within $\{f_i\}$ into $\{g_i\}$ using the unpaired intersection points between $\{f_i\}$ and $\{g_i\}$ (see Figure 16.2). This replaces $\{f_i\}$ by a collection of discs, which we still call $\{f_i\}$ and with the same framed boundaries satisfying $\lambda(f_i, f_j) = \mu(f_i) = 0$ for all i, j. Moreover, we still have that $\lambda(f_i, g_j) = \delta_{ij}$. Apply Theorem 20.4 to achieve the conclusion of the disc embedding theorem. □

We can now prove the sphere embedding theorem, stated next, which we will apply in Chapter 22 to prove the exactness of the surgery sequence. The key difference from Theorem 20.4 is that we embed spheres instead of discs.

SPHERE EMBEDDING THEOREM. *Let M be a smooth, connected 4-manifold such that $\pi_1(M)$ is good. Suppose there exists an immersed collection*

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

of spheres with $\lambda(f_i, f_j) = 0$ for every $i \neq j$ and $\mu(f_i) = 0$ for all i. Suppose moreover that there is an immersed collection

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

of framed dual spheres, that is $\lambda(f_i, g_j) = \delta_{ij}$ for all i, j.

Then there exists an embedding

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

of a collection of spheres in M, with each $\overline{f}_i$ regularly homotopic to f_i, together with framed geometrically transverse spheres,

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

with $\overline{g}_i$ homotopic to g_i for each i.

The sphere embedding theorem is summarised in Figure 20.5. Note that the assumption $\mu(f_i) = 0$ implies that all the self-intersections of f_i can be paired up with Whitney discs, but it does not imply that f_i has trivial normal bundle. Since $\overline{f}_i$ is regularly homotopic to f_i, and f_i has a normal bundle, albeit not a trivial one, we deduce that so does $\overline{f}_i$. Moreover the Euler numbers of the normal bundles of f_i and $\overline{f}_i$ coincide.

PROOF. For each i, find a point on f_i away from all intersections and self-intersections of $\{f_i\}$. Choose a small open ball around this point. Use embedded 1-handles in M disjoint from $\bigcup\{f_i\} \cup \bigcup\{g_i\}$ to connect these small balls into one large open ball B. Let $N := M \smallsetminus B$. Since $\pi_1(N) \cong \pi_1(M)$ and removing B does not change any intersection and self-intersection numbers, we may apply Theorem 20.4 to N and the discs $\{f_i \smallsetminus (f_i \cap B)\}$. This replaces the discs by regularly homotopic, disjointly embedded, flat discs equipped with an immersed collection of framed, geometrically transverse spheres in N. Gluing together B and N as well as $\{f_i \cap B\}$ and the embedded discs just constructed produces the desired embedded spheres $\{\overline{f}_i\}$ in M. □

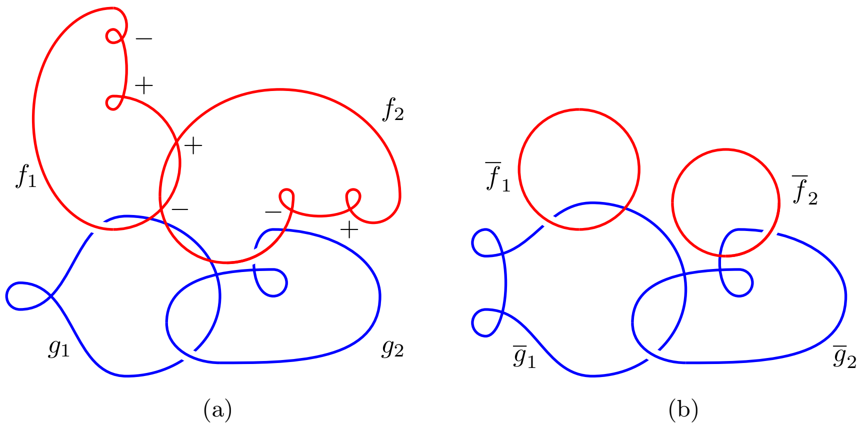


Figure 20.5. Summary of the sphere embedding theorem. We start with the situation in (a) and produce the situation in (b).

CHAPTER 21

The development of topological 4-manifold theory

Mark Powell and Arunima Ray

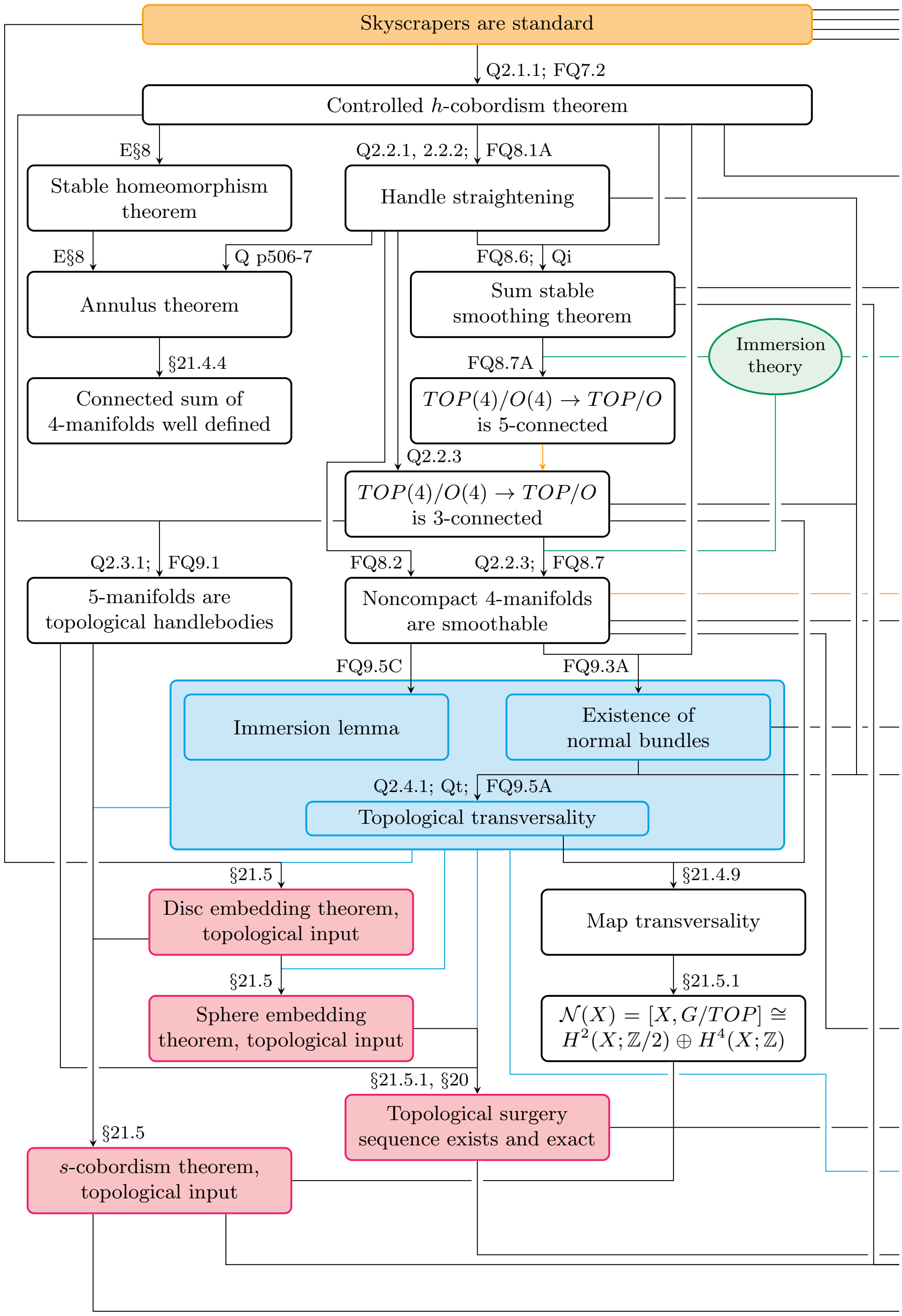
We present a flowchart (Figure 21.1), showing the interdependence among many key statements in the development of topological 4-manifold theory. We explain the statements associated with each node in Figure 21.1 and give references for the implications. Where appropriate we give short outlines of proofs, pointing out the necessary ingredients. In general these explanations are rather limited, aiming to give guidance to further literature, but not to rework it.

As already discussed in Chapter 1, we only prove a category losing disc embedding theorem in this book. That is, we begin with immersed discs in a smooth ambient 4-manifold but only produce flat embedded discs in the outcome. However, with extra work, the hypothesis of a smooth manifold can be weakened to the hypothesis of a topological manifold. The proof of this category preserving disc embedding theorem uses tools, largely developed by Quinn, that depend on the category losing disc embedding theorem.

While the full proof of the category preserving disc embedding theorem is beyond the scope of this book, we indicate the strategy here. The principle is to run the same proof as described in Parts II and IV, but in a topological ambient manifold. In order to do so, one needs, for instance, the fact that locally flat submanifolds in a merely topological ambient 4-manifold may be made transverse by a small isotopy, and the *immersion lemma*, stating that continuous maps of discs and surfaces in a topological 4-manifold can be approximated by generic immersions. The former fact was proven by Quinn (Theorem 1.5) as a consequence of the existence of normal bundles for locally flat submanifolds (Theorem 1.4), a result that also uses Quinn's 5-dimensional controlled h-cobordism theorem, which itself depends on the fact that skyscrapers are standard, as proven in this book.

One starts to see that the logical structure of topological 4-manifold theory is somewhat complicated. Tracking this interdependence in the literature can be challenging. To aid exploration of topological 4-manifolds beyond this book, we have organised the implications into the flowchart shown in Figure 21.1. It should be clear that the facts proven in this book, such as the category losing disc embedding theorem and the fact that skyscrapers are standard, are just the beginning. On the other hand, the entire subsequent development rests on the constructions in this book, whence the high level of detail we provide.

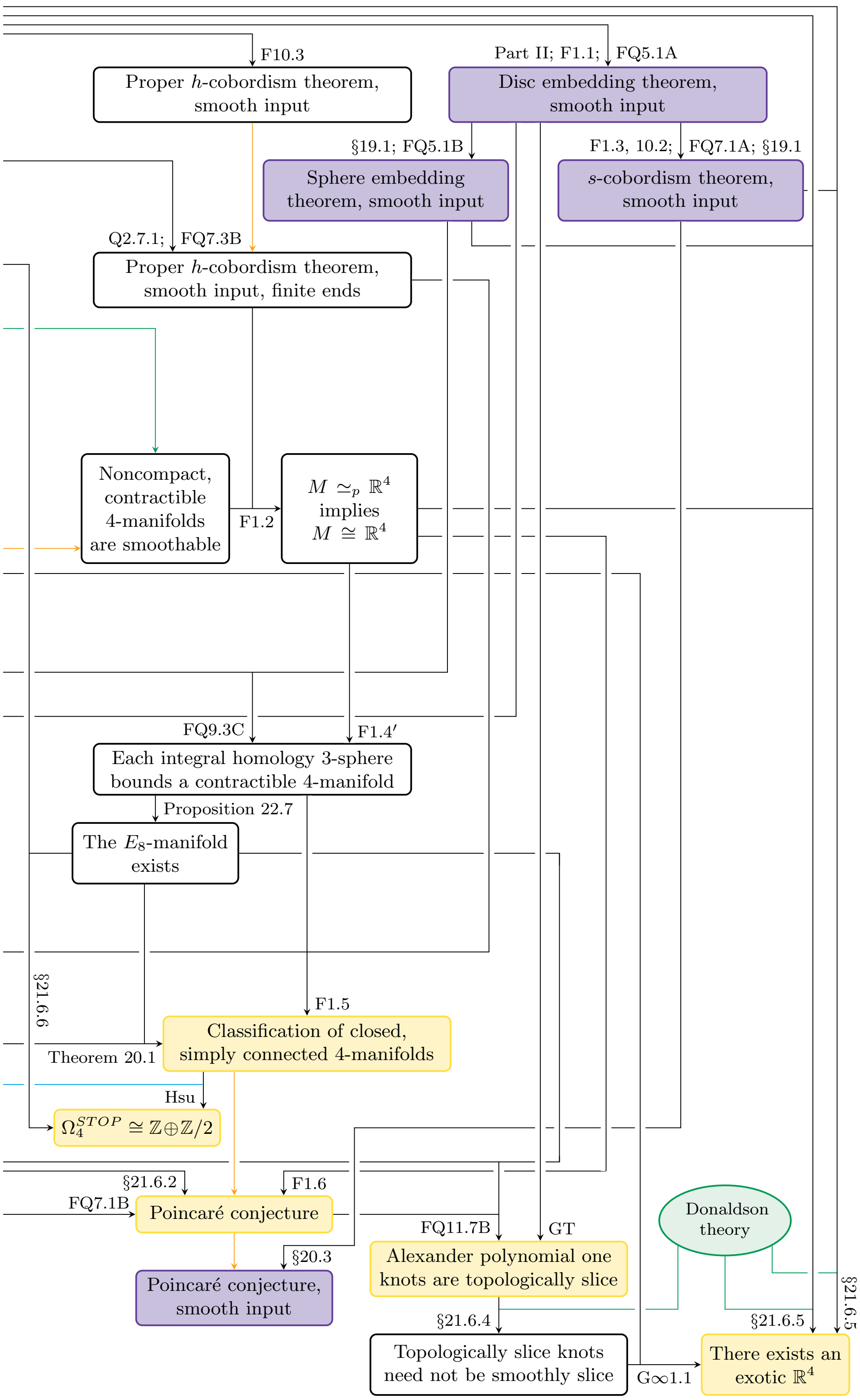
In the flowchart, purple boxes denote the results proven in this book. Green circles indicate results imported from the smooth theory, namely immersion theory and Donaldson's theorems. The blue boxes indicate fundamental results central to constructions in topological 4-manifolds. Yellow boxes denote the flagship results of the field, such as the Poincaré conjecture and the classification of closed, simply connected 4-manifolds up to homeomorphism. The flowchart describes multiple possible routes to each. Orange arrows indicate direct corollaries.



Key

F: [Fre82a] Q: [Qui82b] E: [Edw84]
Ficm: [Fre84] Qt: [Qui88] GT: [GT04]
FQ: [FQ90] Qi: [Qui86] G∞: [Gom85]
Hsu: [Hsu87]

Figure 21.1. The development of topological 4-manifold theory.



F10.3
Part II; F1.1; FQ5.1A
Proper h-cobordism theorem, smooth input
Disc embedding theorem, smooth input
§19.1; FQ5.1B
F1.3, 10.2; FQ7.1A; §19.1
Sphere embedding theorem, smooth input
s-cobordism theorem, smooth input
Q2.7.1; FQ7.3B
Proper h-cobordism theorem, smooth input, finite ends
Noncompact, contractible 4-manifolds are smoothable
F1.2
$M \simeq_p \mathbb{R}^4$ implies $M \cong \mathbb{R}^4$
FQ9.3C
F1.4′
Each integral homology 3-sphere bounds a contractible 4-manifold
Proposition 22.7
The E_8-manifold exists
§21.6.6
F1.5
Theorem 20.1
Classification of closed, simply connected 4-manifolds
Hsu
$\Omega_4^{STOP} \cong \mathbb{Z} \oplus \mathbb{Z}/2$
§21.6.2
F1.6
FQ7.1B
Poincaré conjecture
Donaldson theory
FQ11.7B
GT
§20.3
Alexander polynomial one knots are topologically slice
Poincaré conjecture, smooth input
§21.6.4
§21.6.5
§21.6.5
Topologically slice knots need not be smoothly slice
G∞1.1
There exists an exotic $\mathbb{R}^4$

21.1. Results proven in this book

The entire development of topological 4-manifold theory rests on the statement that skyscrapers are standard. This is one of the main theorems of this book, stated in Theorem 27.1.

THEOREM (Skyscrapers are standard). *For every skyscraper $\widehat{\mathcal{S}}$ there is a homeomorphism of pairs*

$$(F, \partial F)\colon \big(\widehat{\mathcal{S}}, \partial_- \widehat{\mathcal{S}}\big) \cong \big(D^2 \times D^2, S^1 \times D^2\big)$$

that is a diffeomorphism on a collar of $\partial_- \widehat{\mathcal{S}}$, such that if $\Phi\colon S^1 \times D^2 \to \partial_- \widehat{\mathcal{S}}$ is the attaching region, then $\partial F \circ \Phi = \mathrm{Id}_{S^1 \times D^2}$.

As mentioned in Chapter 1, the original work of Freedman was in terms of *Casson handles* rather than skyscrapers. Casson handles are constructed by gluing together neighbourhoods of immersed discs obtained as traces of null homotopies, and as such, the construction relies heavily on the ambient manifold being simply connected. Skyscrapers, on the other hand, may be constructed in an ambient 4-manifold with good, not necessarily trivial, fundamental group, as shown in Part II. Moreover, capped towers may be contracted to produce Casson towers and conversely, any Casson tower contains a capped grope with the same attaching region [**Ray15**, Proposition 3.1]. Thus, one may translate back and forth between the two types of objects. However, as mentioned previously, the decomposition space theory arguments needed for Casson handles is more involved than for skyscrapers. These two points explain our choice to focus on skyscrapers in this book, rather than Casson handles. The proofs using Casson towers and Casson handles may be replaced with the proofs given in this book, or those in [**FQ90**], using capped gropes and skyscrapers. Indeed, throughout the development, one can now replace the use of Casson handles with skyscrapers.

Using the fact that skyscrapers are standard, we deduce the disc embedding theorem with smooth input. This was stated in Chapter 11. The proof modulo the fact that skyscrapers are standard is given in Section 18.4.

THEOREM (Disc embedding theorem, smooth input). *Let M be a connected, smooth 4-manifold, with nonempty boundary and with $\pi_1(M)$ a good group. Let $f_1, \dots, f_n\colon (D^2, S^1) \looparrowright (M, \partial M)$ be a properly immersed collection of discs in M with pairwise disjoint, embedded boundaries. Suppose there is an immersed collection $g_1, \dots, g_n\colon S^2 \looparrowright M$ of framed dual spheres, 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 exist pairwise disjoint, flat, embedded discs $\overline{f}_1, \dots, \overline{f}_n\colon D^2 \hookrightarrow M$, with an immersed collection of framed geometrically transverse spheres $\overline{g}_1, \dots, \overline{g}_n\colon S^2 \looparrowright M$, such that for every i the discs $\overline{f}_i$ and f_i have the same framed boundary and the sphere $\overline{g}_i$ is homotopic to g_i.*

Remarkably, the disc embedding theorem with smooth input in a simply connected ambient 4-manifold, which is the theorem originally proved by Freedman in [**Fre82a**] using Casson handles, is sufficient for much of the subsequent work shown in Figure 21.1.

The above theorem was applied in Chapter 20 to deduce the s-cobordism theorem (Theorem 20.1) and the sphere embedding theorem (Theorem 20.4), both with smooth input and topological output.

THEOREM (Compact s-cobordism theorem, smooth input). *Let N be a smooth, compact 5-dimensional h-cobordism between closed 4-manifolds M_0 and M_1 with vanishing Whitehead torsion $\tau(N, M_0)$. Further, suppose that $\pi_1(N)$ is a good group. Then N is homeomorphic to the product $M_0 \times [0,1]$. In particular, M_0 and M_1 are homeomorphic.*

Theorem (Sphere embedding theorem, smooth input). *Let M be a smooth, connected 4-manifold such that $\pi_1(M)$ is good. Suppose there exists an immersed collection $f_1, \dots, f_n \colon S^2 \looparrowright M$, with $\lambda(f_i, f_j) = 0$ for every $i \neq j$ and $\mu(f_i) = 0$ for $i = 1, \dots, n$. Suppose moreover that there is an immersed collection $g_1, \dots, g_n \colon S^2 \looparrowright M$ of framed dual spheres, that is $\lambda(f_i, g_j) = \delta_{ij}$ for all $i, j = 1, \dots, n$. Then there exist pairwise disjoint, locally flat, embedded spheres $\overline{f}_1, \dots, \overline{f}_n \colon S^2 \hookrightarrow M$, with each $\overline{f}_i$ regularly homotopic to f_i, and with an immersed collection of framed geometrically transverse spheres $\overline{g}_1, \dots, \overline{g}_n \colon S^2 \looparrowright M$, with each $\overline{g}_i$ homotopic to g_i.*

The s-cobordism theorem with smooth input implies the smooth input Poincaré conjecture (Theorem 20.3).

Theorem (Poincaré conjecture, smooth input). *Every closed, smooth 4-manifold homotopy equivalent to the 4-sphere S^4 is homeomorphic to S^4.*

Observe that as a consequence of the category preserving Poincaré conjecture it is known that any homotopy 4-sphere is smoothable. However, this is not evident *a priori* and so cannot be used yet.

21.2. Input to the flowchart

The green ovals in Figure 21.1 show results that are imported in from outside the realm of topological manifolds. While there are in reality many such results, we highlight immersion theory and Donaldson's gauge theory as having been particularly influential.

21.2.1. Immersion theory and smoothing noncompact, contractible 4-manifolds. The concept of a microbundle will appear several times during this chapter, so we recall the definition.

Definition 21.1. An n-dimensional *microbundle* ξ consists of a base space B and a total space E sitting in a diagram

$$B \xrightarrow{i} E \xrightarrow{r} B,$$

such that $r \circ i = \mathrm{Id}_B$, and that is *locally trivial* in the following sense: for every point $b \in B$, there exists an open neighbourhood U, an open neighbourhood V of $i(b)$ and a homeomorphism $\phi_b \colon V \to U \times \mathbb{R}^n$ such that

$$\begin{tikzcd}
 & V \arrow[dr, "r"] \arrow[dd, "\phi_b", "\cong"'] & \\
U \arrow[ur, "i"] \arrow[dr, "\times\{0\}"'] & & U \\
 & U \times \mathbb{R}^n \arrow[ur, "pr_1"'] &
\end{tikzcd}$$

commutes.

Immersion theory, due to Lashof [Las70a, Las70b, Las70c, Las71], implies the following key fact about noncompact 4-manifolds. Below $BTOP(4)$ and $BO(4)$ denote the classifying spaces for 4-dimensional topological microbundles and rank 4 smooth vector bundles, respectively. By the Kister-Mazur theorem [Kis64, Theorem 2], there is a natural correspondence between n-dimensional microbundles and fibre bundles with fibre $\mathbb{R}^n$ and structure group $TOP(n)$, the group of homeomorphisms of $\mathbb{R}^n$ that preserve the origin. We therefore use the concepts of 4-dimensional microbundles and fibre bundles with $\mathbb{R}^4$ fibre interchangeably. The *tangent microbundle* of a 4-manifold, given by $M \xrightarrow{\Delta} M \times M \xrightarrow{r} M$, where

$\Delta(m) = (m, m)$ is the diagonal map and $r(m, n) = m$, corresponds via the Kister-Mazur theorem to a map $\tau_M\colon M \to BTOP(4)$. This is also called the *topological tangent bundle* of M.

THEOREM. *Let M be a connected, noncompact 4-manifold and let $\tau'_M\colon M \to BO(4)$ be a lift of the tangent microbundle classifying map $\tau_M\colon M \to BTOP(4)$ along the forgetful map $BO(4) \to BTOP(4)$. Then M admits a smooth structure whose tangent bundle classifying map is homotopic to τ'_M.*

This implies the following smoothing result, since there is no obstruction to lifting for a contractible space. Recall that a map between spaces $f\colon X \to Y$ is called *proper* if the inverse image $f^{-1}(K)$ of every compact set $K \subseteq Y$ is compact in X. A proper map $f\colon X \to Y$ is said to be a *proper homotopy equivalence* if there is a proper map $g\colon Y \to X$ such that $f \circ g$ and $g \circ f$ are properly homotopic to Id_Y and Id_X respectively, meaning that the homotopies are proper maps.

THEOREM (Contractible 4-manifolds are smoothable). *If M is a connected, noncompact, contractible 4-manifold, such as a manifold proper homotopy equivalent to $\mathbb{R}^4$, then M admits a smooth structure.*

21.2.2. Donaldson theory. In 1983, Donaldson applied techniques from Yang-Mills gauge theory to establish strong restrictions on the intersection forms of smooth 4-manifolds. That such restrictions exist had been known since 1952 by Rochlin's theorem [**Roc52**], which states that the intersection form of a smooth, closed, spin 4-manifold must have signature divisible by 16. By studying *instantons*, that is anti-self-dual connections on $SU(2)$-bundles on definite, smooth 4-manifolds, Donaldson proved the following.

THEOREM (Donaldson's Theorem A). *Let M be a closed, smooth, oriented 4-manifold with negative definite intersection form. Then the form is equivalent over the integers to the standard intersection form $\oplus[-1]$.*

This theorem was proved in [**Don83**] with the restriction that M be simply connected, but this condition was later removed in [**Don87b**]. Donaldson gave further restrictions on intersection forms of smooth 4-manifolds in [**Don86**]. Alternative proofs were later given using Seiberg-Witten gauge theory (see e.g. [**GS65**, Section 2.4.2]). Seiberg-Witten theory also leads to further restrictions on intersection forms of closed 4-manifolds, such as Furuta's 10/8 theorem [**Fur01**], recently extended in [**HLSX18**].

The results of Donaldson and Freedman combine to exhibit a remarkable disparity between the smooth and topological categories in dimension 4. First, in contrast to Donaldson's Theorem A every definite, nonsingular, symmetric, bilinear, integral form is realised as the intersection form of some closed topological 4-manifold. The intersection form also does not characterise simply connected smooth 4-manifolds. For example, in [**Don87a**], Donaldson showed the following.

THEOREM (Smooth failure of the h-cobordism theorem). *The closed 4-manifold $\mathbb{CP}^2\#9\overline{\mathbb{CP}^2}$ and its $(2,3)$ logarithmic transform L, called the Dolgachev surface, are homeomorphic but not diffeomorphic.*

Wall [**Wal64**] showed that any two closed, smooth, simply connected 4-manifolds with isomorphic intersection forms are smoothly h-cobordant. Consequently, the above two smooth 4-manifolds are smoothly h-cobordant. By the h-cobordism theorem with smooth input (Theorem 20.2), such an h-cobordism is topologically a product. Since the manifolds above are not diffeomorphic, the h-cobordism is not diffeomorphic to a product. So there is no smooth 5-dimensional s-cobordism theorem.

We will see in Section 21.6.4 how to apply Donaldson's results to show that not all topologically slice knots are smoothly slice, and that $\mathbb{R}^4$ has exotic smooth structures.

21.3. Further results from Freedman's original paper

We now consider the main results from [**Fre82a**] that we do not prove in this book. As seen in the flowchart (Figure 21.1), there are alternative routes to many of these results via the work of Quinn [**Qui82b**], which we describe in a later section.

21.3.1. The proper h-cobordism theorem with smooth input. The h-cobordism theorem is a powerful tool to show that two manifolds are homeomorphic. However, the version of the h-cobordism theorem mentioned earlier (Theorem 20.1), first proved in [**Fre82a**], begins with a compact, *smooth* h-cobordism and concludes that it is homeomorphic to a product. Smoothness is essential in the proof, which consists of modifying a handle decomposition of the cobordism. One cannot prove the category preserving Poincaré conjecture, that any closed, *topological* 4-manifold homotopy equivalent to S^4 is homeomorphic to S^4, using the version of the h-cobordism theorem in Theorem 20.3, without first knowing that every homotopy S^4 is smoothable.

However, as we saw previously, immersion theory may be applied to deduce that noncompact, contractible 4-manifolds are smoothable. We then need a noncompact version of the h-cobordism theorem. One such result was proved in [**Fre82a**]. Recall that a *proper h-cobordism* is a cobordism N between manifolds M_0 and M_1 such that the inclusions $M_i \hookrightarrow N$ are proper homotopy equivalences.

Theorem (Proper h-cobordism theorem, smooth input). *Let N be a smooth, simply connected, 5-dimensional proper h-cobordism between 4-manifolds M_0 and M_1. Assume further that N is simply connected at infinity. Then N is homeomorphic, via a proper map, to the product $M_0 \times [0,1]$. In particular, M_0 and M_1 are homeomorphic.*

The definition of simply connectivity at infinity was given in Definition 7.3. Note that the fundamental groups in question, of the space as well as the end, are required to be trivial in the statement above.

Freedman's proof is based on Siebenmann's sketch of a proof for the corresponding high dimensional result [**Sie70**]. Siebenmann's idea is to partition the given proper h-cobordism N into a union of compact h-cobordisms. In the high dimensional proof, the Whitney move is required to construct these h-cobordisms, and then the compact h-cobordism theorem is applied to each constituent h-cobordism. In the 5-dimensional case, this approach does not work directly, since performing the Whitney move loses category, as the disc embedding theorem only produces flat Whitney discs. Consequently the result of the Whitney move is a sequence of topological h-cobordisms. Freedman did not have access to the fully topological h-cobordism theorem when writing [**Fre82a**]. Instead, he used an intricate argument to carefully arrange for Casson handles pairing the extraneous intersections between belt spheres of 2-handles and attaching spheres of 3-handles to perform a "grand cancellation" at the end of the proof. The availability of category preserving results should make this proof simpler, although this has not been written down.

While the theorem as stated above has no restrictions on the number of ends of N, the special case that N has only finitely many ends can be deduced from the controlled h-cobordism theorem, as shown in [**Qui82b**, Theorem 2.7.1; **FQ90**, Corollary 7.3B]. We will discuss the controlled h-cobordism theorem in Section 21.4.1.

The version of the proper h-cobordism theorem, with finitely many ends, in [**FQ90**, Corollary 7.3B] is expanded to include good fundamental groups, provided the ends also have good local fundamental groups. The version with infinitely many ends ought to admit an extension to good fundamental groups as well, by replacing Casson handles with skyscrapers in Freedman's proof, but such a proof has also not yet been written down.

The special case that N has a single end forms an important component of the classification result for closed, simply connected 4-manifolds given in [**Fre82a**]. However, that result was at that time limited to 4-manifolds smoothable away from a point and requires work of Quinn [**Qui82b**], that every connected 4-manifold is smoothable away from a point, to be promoted to a full classification, as we see in Section 21.6.1. The single end case also yields the following corollary [**Fre82a**, Corollary 1.2], which is our first category preserving statement.

Theorem ($M \simeq_p \mathbb{R}^4$ implies $M \cong \mathbb{R}^4$). *Let M be a topological 4-manifold with empty boundary proper homotopy equivalent to $\mathbb{R}^4$, denoted $M \simeq_p \mathbb{R}^4$. Then M is homeomorphic to $\mathbb{R}^4$.*

It is now known that for $n \geq 3$, an open, contractible n-manifold M is homeomorphic to $\mathbb{R}^n$ if and only if M is simply connected at infinity. For $n \geq 5$ this is due to Stallings [**Sta62**]. Stallings' result applies to PL-manifolds, but we can use immersion theory, which was described in dimension four in Section 21.2.1 but which holds in all dimensions at least four, to find a PL structure on M. In dimension three, the result was shown by Edwards and Wall [**Edw63**, **Wal65b**] modulo the 3-dimensional Poincaré conjecture, which was settled by Perelman [**Per02**, **Per03b**, **Per03a**] (see also [**MT07**, **KL08**]). There exist open, contractible manifolds that are not simply connected at infinity, such as the Whitehead manifold from Chapter 7.

Proof. Since M is proper homotopy equivalent to $\mathbb{R}^4$, M is noncompact and contractible. Immersion theory (Section 21.2.1) implies that M has a smooth structure. Let M_{sm} denote M equipped with some smooth structure and let $B \subseteq M_{\mathrm{sm}}$ be the interior of a smooth ball in M_{sm}. Then $M_{\mathrm{sm}} \times [0,1) \cup B \times \{1\}$ is a smooth, proper h-cobordism between $M_{\mathrm{sm}} \times \{0\}$ and $B \times \{1\}$, that is simply connected as well as simply connected at infinity with a single end. It follows from the proper h-cobordism theorem that $M \cong \mathbb{R}^4$. $\square$

This result yields the fully topological Poincaré conjecture [**Fre82a**], as we sketch in Section 21.6.2. Since the result enables us to recognise 4-dimensional Euclidean space $\mathbb{R}^4$, at least up to homeomorphism, it also appears in proofs of the existence of exotic smooth structures on $\mathbb{R}^4$, as we shall see in Section 21.6.5.

21.3.2. Integral homology 3-spheres bound contractible 4-manifolds. An integral homology 3-sphere is a closed 3-manifold whose homology groups with integer coefficients agree with those of S^3.

Theorem (Every homology sphere bounds a contractible 4-manifold). *Every integral homology 3-sphere Σ is the boundary of some compact, contractible, topological 4-manifold.*

Freedman's paper [**Fre79**] was the first major application of Casson's constructions to the study of topological 4-manifolds. In it, Freedman showed that surgery theory works in smooth, simply connected 4-manifolds at the expense of removing isolated points from both the domain and codomain. A consequence [**Fre79**, Theorem 4(C)] was that for every integral homology 3-sphere Σ there is a smooth, proper embedding of the punctured manifold $\Sigma \smallsetminus \{*\}$ in an open, smooth 4-manifold proper homotopy equivalent to $\mathbb{R}^4$. The theorem above, which was [**Fre82a**, Theorem 1.4′], is a consequence of the previous section together with a result of Kirby that other than surfaces in 3-manifolds, no codimension one embedding can have an isolated non-flat point [**Kir68**, Theorem 1].

See [**Ker69**, Theorem 3; **HH67**, Theorem 5.6] for higher-dimensional analogues of the above statement.

An alternative proof was given by Freedman-Quinn in [**FQ90**, Corollary 9.3C] (see Proposition 22.6). The argument there uses the sphere embedding theorem with smooth input to improve $\Sigma \times [0,1]$ to a topological, simply connected 4-manifold V with $\partial V = \Sigma \sqcup -\Sigma$ and $H_2(V) = 0$. Then stack together infinitely many copies of V and compactify the result with a single point. That this last point is a manifold point uses [**Qui82b**, Theorem 2.5.1; **FQ90**, Theorem 9.3A] on normal bundles for locally flat submanifolds in a 4-manifold. We discuss the latter result in Section 21.4.8.

Remark 21.2. For a fixed homology 3-sphere Σ, the contractible 4-manifold constructed above is unique up to homeomorphism relative to the boundary. This requires further ingredients. Here is a sketch of the proof. Let W and W' be two contractible 4-manifolds with boundary a homology 3-sphere Σ. By the topological input Poincaré conjecture (Section 21.6.2), the union $W \cup_\Sigma -W'$ is homeomorphic to S^4, so bounds a 5-ball $V = D^5$. Decompose the boundary as $W \cup \Sigma \times [0,1] \cup -W'$ to view V as a simply connected h-cobordism relative boundary. The category preserving compact h-cobordism theorem (Section 21.5) implies that V is homeomorphic to $W \times [0,1]$ and thus W is homeomorphic to W' relative to the boundary.

The theorem that every integral homology 3-sphere bounds a contractible 4-manifold is used to prove the following [**Fre82a**, Theorem 1.7].

Theorem (The E_8-manifold exists). *There exists a closed, topological, simply connected 4-manifold whose intersection form is isometric to the form $\lambda\colon \mathbb{Z}^8 \times \mathbb{Z}^8 \to \mathbb{Z}$ represented by the E_8 matrix.*

The proof is straightforward (see Proposition 22.7): plumb together D^2-bundles over S^2 with Euler number 2 according to the E_8 matrix to produce a smooth, simply connected 4-manifold with integral homology sphere boundary and then cap off with the topological, contractible 4-manifold given by the previous result. The same strategy was used by Freedman to construct closed, topological, simply connected 4-manifolds realising any given nonsingular, symmetric, integral bilinear form as its intersection form, as we describe in Section 21.6.1. As noted in Chapter 1, Rochlin's theorem implies that there is no closed, smooth 4-manifold realising the E_8 matrix as its intersection form.

We single out the case of the E_8-manifold because its existence is crucial to many later developments. In particular it is used in the classification of closed, simply connected 4-manifolds, up to homeomorphism, using surgery, as we describe in Section 21.6.1 and in further detail in Chapter 22, as well as the computation of the oriented bordism group Ω_4^{STOP} (Section 21.6.6). It appears to be a necessary ingredient in any effective application of the surgery sequence in dimension four. It is not known how to deduce the existence of the E_8-manifold from the surgery sequence without first having an independent construction of the E_8-manifold. In particular, for X a simply connected Poincaré complex with intersection form E_8, why, *a priori*, should the map from $\mathcal{N}(X) \to L_4(\mathbb{Z})$ be onto? This appears to be an omission in the argument given for the construction of contractible 4-manifolds in [**FQ90**, Section 11.4].

21.4. Foundational results due to Quinn

Some of the most dramatic results in [**Fre82a**], such as the Poincaré conjectue and the classification of closed, simply connected 4-manifolds (smoothable away from a point), were derived from the proper h-cobordism theorem. That is, the extension to noncompact settings was extremely useful. In [**Qui82b**], Quinn proved a second noncompact version of the h-cobordism theorem, namely the *controlled h-cobordism theorem* (sometimes called the *thin h-cobordism theorem* or *ε-h-cobordism theorem.* This result formed the crucial input for a swathe of further developments which were foundational to the theory of topological 4-manifolds, and are shown on the left of the flowchart (Figure 21.1).

21.4.1. The controlled h-cobordism theorem with smooth input. Quinn proved the controlled h-cobordism theorem in [**Qui82b**, Theorem 2.1.1] (see also [**FQ90**, Chapter 7]). The proof of the simplest case of the theorem, namely the case of cobordisms homeomorphic to $\mathbb{R}^4 \times [0,1]$, which is sufficient to prove the annulus theorem, is described in considerable detail in [**Edw84**]. We recommend that the reader wishing to understand Quinn's proof in depth start with [**Edw84**] to get a sense of the intricacy of the argument.

Now we now introduce the terminology necessary to understand the statement of the controlled h-cobordism theorem. Let X be a metric space and let $\delta\colon X \to (0,\infty)$ be a continuous function. In the sequel, the space X will serve as a reference space. Let N be a smooth 5-dimensional h-cobordism between potentially noncompact smooth 4-manifolds M_0 and M_1, equipped with a proper map $F\colon N \to X$. Since N is an h-cobordism, N deformation retracts to M_0 and to M_1. That is, for $i = 0,1$ there are maps $g_i\colon N \to M_i$ with the composition $M_i \xrightarrow{j_i} N \xrightarrow{g_i} M_i$ the identity, and a homotopy $G\colon j_i \circ g_i \sim \mathrm{Id}_N\colon N \times [0,1] \to N$, where $j_i\colon M_i \hookrightarrow N$ are the inclusion maps.

Suppose that for every $n \in N$ the track of n under this homotopy, namely $F \circ G(\{n\} \times [0,1]) \subseteq X$, lies in the ball in X of radius $\delta(F \circ G(\{n\} \times \{0\}))$ with centre $F \circ G(\{n\} \times \{0\})$. Then we say that the homotopy G has *diameter less than* δ, and that N is a *(δ, h)-cobordism.*

A map $F\colon N \to X$ is said to be *$(\delta,1)$-connected* if, for any relative 2-dimensional CW complex (K, L), the diagram

$$\begin{tikzcd} L \arrow[r] \arrow[d] & N \arrow[d, "F"] \\ K \arrow[r, "p"'] \arrow[ur, dashed, "q"] & X \end{tikzcd}$$

has a solution, meaning that there is a map $q\colon K \to N$ such that the top triangle commutes and the failure of the bottom triangle to commute is bounded by δ. This means that $F \circ q(x)$ and $p(x)$ are distance less than $\delta(p(x))$ apart, as measured in the metric on X for all $x \in K$. It is instructive to think of the case where (K, L) is a 2-disc relative to its boundary. The condition of $(\delta,1)$-connectedness allows us to find discs in N (usually Whitney discs) of controlled diameter.

Like other versions of the h-cobordism theorem, the controlled h-cobordism theorem will show that certain h-cobordisms are products. In addition, the product structure will be controlled by a given function. This is made precise in the following definition. Let $\varepsilon\colon X \to (0,\infty)$ be a continuous function. For $P\colon M_0 \times [0,1] \to N$ a homeomorphism demonstrating that N has a product structure, the composition $K_0 := F \circ P\colon M_0 \times [0,1] \to N \to X$ is a homotopy, as is $K_1 := F \circ P \circ (\mathrm{Id} \times \mathrm{rev})\colon M_0 \times [0,1] \to X$, where $\mathrm{rev}\colon [0,1] \to [0,1]$ sends $t \mapsto 1-t$. If both the homotopies K_0 and K_1 satisfy that for every $m \in M_0$, the track $K_i(\{m\} \times [0,1])$ of

m lies in the ball in X of radius $\varepsilon(F \circ K_i(\{m\} \times \{0\}))$ with centre $F \circ K_i(\{m\} \times \{0\})$, then P is said to be an *ε-product structure.*

THEOREM (Controlled h-cobordism theorem, smooth input). *Let X be a locally compact, locally 1-connected metric space and let $\varepsilon\colon X \to (0, \infty)$ be a continuous function. Then there exists a continuous function $\delta\colon X \to (0, \infty)$ such that every smooth 5-dimensional $(\delta, 1)$-connected (δ, h)-cobordism N between smooth 4-manifolds M_0 and M_1, equipped with a proper map $F\colon N \to X$, has an ε-product structure.*

A more general version of the above statement exists [**FQ90**, Theorem 7.2C], where X is only required to have good local fundamental groups, rather than be locally 1-connected. In this case one requires the vanishing of the *controlled torsion* of N [**Qui82a**], which is a controlled, noncompact analogue of Whitehead torsion.

We briefly sketch the proof of the controlled h-cobordism theorem, which follows the same basic strategy as the proof of the compact s-cobordism theorem. Given a handlebody decomposition of N, perform handle cancellation and trading to arrange that there are only 2- and 3-handles. Extraneous intersection points are paired by immersed Whitney discs. As in the compact case, we wish to replace them by flat, mutually disjoint and embedded Whitney discs. Whitney moves using these discs would then allow us to cancel the 2- and 3-handles. However, unlike the proof of the compact s-cobordism theorem, we cannot use the disc embedding theorem directly since if N is noncompact, we need to embed and separate infinitely many immersed Whitney discs. Quinn's approach was to prove a noncompact, controlled version of the disc embedding theorem, which he calls the *disc deployment lemma* [**Qui82b**, Lemma 3.2; **FQ90**, Theorem 5.4]. This shows how to replace an infinite family of discs, equipped with transverse spheres and with sizes controlled by a given function, by mutually disjoint, flat and embedded discs. The strategy involves carefully grouping the given immersed discs into collections, guided by the controlling function, and then working within each such collection to build mutually disjoint skyscrapers of controlled size, pairing all the intersections.

Applying the fact that skyscrapers are standard gives mutually disjoint and embedded Whitney discs, Whitney moves over which yields the desired embedded discs. Apply the disc deployment lemma to the middle level of the cobordism N to complete the proof of the theorem.

Each step of the proof increases the size of the relevant data in a bounded manner (similar to our analysis in the proof of Theorem 19.4). The $(\delta, 1)$-connected hypothesis is needed whenever any discs are produced via a null homotopy, to ensure that the sizes are not too large. Finally, by choosing the initial handle decomposition to be small enough, which determines the function δ, one ensures that the final product structure is bounded by ε.

As mentioned in Section 21.3.1, the proper h-cobordism theorem in the case of finitely many ends can be derived as a corollary of the controlled h-cobordism theorem, as shown in [**Qui82b**, Theorem 2.7.1; **FQ90**, Corollary 7.3B]. The proof begins by properly embedding a graph within the given proper h-cobordism. The graph performs the rôle of the metric space X and the cobordism is seen to be a $(\delta, 1)$-connected (δ, h)-cobordism for any given choice of δ after a suitable compression of the graph.

In the remainder of this section, we describe many other implications of the controlled h-cobordism theorem.

21.4.2. Handle straightening. The *handle straightening theorem*, or *generalised annulus conjecture*, appears as [**Qui82b**, Theorem 2.2.2]. A mild generalisation was then given in [**FQ90**, Theorem 8.1]. We give the statement from [**Qui82b**], which attempts to smooth the cores of handles.

THEOREM (Handle straightening). *Fix $j \in \{0,1,2\}$. Let $h\colon D^j \times \mathbb{R}^{4-j} \to W$ be a homeomorphism of smooth manifolds, such that $h|_{S^{j-1}\times\mathbb{R}^{4-j}}$ is smooth. Fix a neighbourhood U of $D^j \times \{0\}$.*

- *(a) If $j = 0,1$, then h is isotopic, relative to the boundary $S^{j-1} \times \mathbb{R}^{4-j}$ and via an isotopy supported inside U, to a homeomorphism $h'\colon D^j \times \mathbb{R}^{4-j} \to W$ that restricts to a diffeomorphism on some neighbourhood $V \subseteq U$ of $D^j \times \{0\}$.*
- *(b) If $j = 2$, then h is isotopic, relative to the boundary $S^1 \times \mathbb{R}^2$ and via an isotopy supported inside U, to a homeomorphism $h'\colon D^2 \times \mathbb{R}^2 \to W$ that restricts to a diffeomorphism on some subset $V \subseteq U \subseteq D^2 \times \mathbb{R}^2$, where V is one of either:*
  - *(i) a neighbourhood of a properly immersed disc in $U \subseteq D^2 \times \mathbb{R}^2$ obtained from $D^2 \times \{0\}$ by a smooth regular homotopy that fixes $S^1 \times \{0\}$; or*
  - *(ii) a neighbourhood of a locally flat, properly embedded disc in $U \subseteq D^2 \times \mathbb{R}^2$ obtained from $D^2 \times \{0\}$ by a topological ambient isotopy that fixes $S^1 \times \{0\}$.*

The proof uses the controlled h-cobordism theorem (Section 21.4.1) applied to the mapping cylinder of the homeomorphism. In fact one needs the *technical controlled h-cobordism theorem* of [**FQ90**, Theorem 7.2C].

21.4.3. The stable homeomorphism theorem. An orientation preserving homeomorphism $f\colon \mathbb{R}^n \to \mathbb{R}^n$ is called *stable* if it can be written as a composition $f_m \circ \cdots \circ f_2 \circ f_1$ where each f_i restricts to the identity on some open set $U_i \subseteq \mathbb{R}^n$.

THEOREM (4-dimensional stable homeomorphism theorem). *Every orientation preserving homeomorphism $f\colon \mathbb{R}^4 \to \mathbb{R}^4$ is stable.*

The version of this theorem with 4 replaced by n is true for all n. In dimension at most three the result is classical [**Rad25**, **Moi52b**]. The high dimensional version for $n \geq 5$, proved by Kirby [**Kir69**] using the *torus trick* in dimension at least 5, was the catalyst for the development of topological manifold theory in these dimensions [**KS77**]. Quinn's proof of the 4-dimensional version [**Qui82b**] was similarly transformative. A key corollary is the *annulus theorem*, which we describe in the next section.

The 4-dimensional stable homeomorphism theorem follows from a simple case of the controlled h-cobordism theorem, which was described in [**Edw84**, Section 8].

21.4.4. The annulus theorem and connected sum. The n-dimensional annulus theorem is a consequence of the n-dimensional stable homeomorphism theorem, for each n. This was shown by Brown and Gluck [**BG64**] prior to Kirby's proof of the stable homeomorphism theorem for $n \geq 5$; see also [**FNOP19**, Section 5] for further details on this deduction.

THEOREM 21.3 (4-dimensional annulus theorem). *Let $f\colon S^3 \to \operatorname{Int} D^4$ be a locally flat embedding. Then the region bounded by $f(S^3)$ and $S^3 = \partial D^4$ is homeomorphic to the annulus $S^3 \times [0,1]$.*

An alternative proof of the annulus theorem, using the handle straightening theorem, was given by Quinn. Since Quinn's argument [**Qui82b**, p. 507] is rather terse, we provide some details.

Proof using handle straightening. Brown's theorem (Theorem 3.11) implies that $f(S^3)$ is bicollared. By the Schoenflies theorem (Theorem 3.29), $f\colon S^3 \to \operatorname{Int} D^4$ extends to a map $F\colon D^4 \to \operatorname{Int} D^4$. Using the collar extend F further to a homeomorphism $\widehat{F}\colon \mathbb{R}^4 \to \operatorname{Int} D^4$ onto a neighbourhood of $F(D^4)$ such that F factors as $F\colon D^4 \hookrightarrow \mathbb{R}^4 \xrightarrow{\widehat{F}} \operatorname{Int} D^4$ via the standard inclusion $D^4 \subseteq \mathbb{R}^4$. Attempting clarity, we refer to the two copies of D^4 as X and Y. That is, we have a map $F\colon X \hookrightarrow \mathbb{R}^4 \xrightarrow{\widehat{F}} \operatorname{Int} Y$.

The handle straightening theorem with $j = 0$ gives an isotopy of $\widehat{F}$, supported on $\operatorname{Int} X$, to a homeomorphism $\widetilde{F}\colon \mathbb{R}^4 \to \operatorname{Int} Y$ such that $\widetilde{F}$ is a diffeomorphism when restricted to some neighbourhood V of 0. By construction, $\widetilde{F}$ equals f on $\partial X = S^3$. Choose a 4-ball $B \subseteq V$. Since ∂B and $\partial X = S^3 \subseteq \mathbb{R}^4$ cobound an annulus in $\mathbb{R}^4$, and $\widetilde{F}$ is a homeomorphism with $\widetilde{F}(\partial X) = f(\partial X)$,

$$\overline{Y \smallsetminus f(X)} \text{ is homeomorphic to } \overline{Y \smallsetminus \widetilde{F}(B)}.$$

But since $\widetilde{F}|_B\colon B \to Y$ is a smooth embedding, by the smooth annulus theorem we have

$$\overline{Y \smallsetminus \widetilde{F}(B)} \text{ is homeomorphic to } S^3 \times [0,1].$$

It follows that $\overline{Y \smallsetminus f(X)}$ is homeomorphic to $S^3 \times [0,1]$ as desired. □

A corollary of the annulus theorem is that the connected sum operation on either nonorientable or oriented topological 4-manifolds is well defined. To show this, one needs to see that any two locally flat embeddings of D^4 in a given connected, topological 4-manifold are isotopic. This is a standard fact (for smooth embeddings) in the smooth category. In the topological category, we may prove this using the annulus theorem, as follows. Given two locally flat embeddings of D^4, connect the centres by an arc covered by coordinate patches, shrink down the image of one of the embeddings until it is contained in a patch, then transport patch by patch to the other endpoint, shrinking further to ensure that the image lies within the second embedding of D^4. This uses the isotopy extension theorem [**EK71**], which uses that our embeddings of D^4 are locally flat and thus have bicollared boundaries by Brown's theorem (Theorem 3.11). Use the annulus theorem to expand the first image onto the second. Finally use the fact that every orientation preserving homeomorphism from S^3 to itself is isotopic to the identity [**Fis60**] (see also [**BZH14**, Section 1.B]) to complete the argument. See [**FNOP19**, Section 5] for more details. To see that the connected sum operation gives another topological manifold one again uses Brown's result (Theorem 3.11). Finally note that the connected sum of unoriented 4-manifolds is not well defined, since for example $\mathbb{CP}^2 \# \overline{\mathbb{CP}^2}$ and $\mathbb{CP}^2 \# \mathbb{CP}^2$ are not homeomorphic.

An alternative proof that connected sum is well defined is also possible using the smooth version and the fact that connected topological 4-manifolds are smoothable away from a point (Section 21.4.7).

21.4.5. The sum stable smoothing theorem. The best-known part of the upcoming *sum stable smoothing theorem* is that for a compact 4-manifold with vanishing *Kirby-Siebenmann invariant*, there is an $n \geq 0$ such that $M \#^n S^2 \times S^2$ admits a smooth structure. The full version of the theorem, from [**FQ90**, Section 8.6], is more precise, linking the smooth structures on the $S^2 \times S^2$-stabilisations to $S^2 \times S^2$-stable lifts of classifying maps for the tangent microbundle, and discussing stable smoothings of homeomorphisms. In order to state the theorem we first recall some bundle theory.

Let TOP be the colimit of the system $TOP(n)$ of homeomorphisms of $\mathbb{R}^n$ that fix the origin, and let $O = \operatorname{colim} O(n)$ be the analogous colimit of the orthogonal groups. Let $BTOP$ and BO be the associated classifying spaces.

The *Kirby-Siebenmann invariant* $\operatorname{ks}(M) \in H^4(M;\mathbb{Z}/2)$ of a closed, connected, topological 4-manifold M is the obstruction for the stable tangent microbundle classifying map $\tau_M\colon M \to BTOP$ to admit a lift $\tau'_M\colon M \to BPL$ along the forgetful map $BPL \to BTOP$. There is a unique obstruction because the homotopy fibre $TOP/PL \simeq K(\mathbb{Z}/2,3)$.

Recall that each smooth structure on a topological 4-manifold N determines a lift $\tau_N\colon N \to BO$ of the classifying map $N \to BTOP$ of the stable tangent microbundle of N, namely the classifying map of the stable tangent vector bundle of N with this choice of smoothing. A smooth structure on a manifold induces a PL structure. In fact for manifolds of dimension at least six, this correspondence induces a bijection between the sets of smooth and PL structures [**HM74**]. In other words, a lift to BPL is equivalent to a lift to BO.

A *sum stabilisation* of a connected 4-manifold M is a connected sum $M\#^n S^2\times S^2$, for some $n \geq 0$. A homeomorphism of connected manifolds $h\colon M \to N$ determines a homomorphism

$$h\#\operatorname{Id}\colon M\#^n S^2 \times S^2 \to N\#^n S^2 \times S^2$$

by removing the ball used for connected sum in the domain, and its image in N, and then extending the restriction of h by the identity on $\#^n S^2 \times S^2 \smallsetminus \mathring{D}^4$.

A smooth structure on $M\#^n S^2\times S^2$ determines a smooth structure on $M\#^{n+1} S^2\times S^2$, by using the standard structure on $S^2 \times S^2$ and that the connected sum operation on smooth manifolds is well defined. A lift

$$\tau_{M\#^n S^2\times S^2}\colon M\#^n S^2 \times S^2 \to BO$$

of the stable tangent microbundle of $M\#^n S^2 \times S^2$ likewise determines a lift

$$\tau_{M\#^{n+1} S^2\times S^2}\colon M\#^{n+1} S^2 \times S^2 \to BO$$

of the stable tangent microbundle of $M\#^{n+1} S^2 \times S^2$, by gluing with the lift $\tau_{S^2\times S^2}\colon S^2\times S^2 \to BO$ corresponding to the standard smooth structure on $S^2\times S^2$. Indeed, the homotopy classes of lifts $\tau_{M\#^n S^2\times S^2}$ are in one to one correspondence with the homotopy classes of lifts $\tau_{M\#^{n+1} S^2\times S^2}$, since $S^2 \times S^2 \smallsetminus \mathring{D}^4 \to BTOP$ comes with a canonical lift.

A smooth structure on $M\#^n S^2 \times S^2$ is said to be $S^2 \times S^2$*-stably equivalent* to a smooth structure on $M\#^{n'} S^2 \times S^2$ if there is an integer $K \geq n, n'$ such that the two induced smooth structures on $M\#^K S^2 \times S^2$ are equivalent, that is, the two smooth manifolds are diffeomorphic. Similarly, we say that two lifts $\tau_{M\#^n S^2\times S^2}$ and $\tau_{M\#^{n'} S^2\times S^2}$ to BO of the respective stable tangent microbundles are $S^2 \times S^2$*-stably homotopic* if there is an integer $K \geq n, n'$ such that the induced lifts $M\#^K S^2 \times S^2 \to BO$ are homotopic. Taking the classifying map of the stable tangent vector bundle, $S^2 \times S^2$-stable equivalence classes of smooth structures map to $S^2 \times S^2$-stable homotopy classes of lifts of stable tangent microbundles. The first part of the upcoming theorem states that this association is in fact a bijection.

Theorem (Sum stable smoothing theorem). *Let M be a connected, topological 4-manifold. There exists a sum stabilisation of M that admits a smooth structure if and only if the Kirby-Siebenmann invariant of M is trivial, which holds if and only if the stable tangent microbundle classifying map $M \to BTOP$ factors as $M \to BO \to BTOP$, where $BO \to BTOP$ is the canonical forgetful map. Moreover, we have the following refinements.*

(1) *Taking the classifying map of the stable tangent microbundle gives rise to a bijection between the set of $S^2 \times S^2$-stable equivalence classes of smooth*

structures on the manifolds $\{M\#^n S^2 \times S^2\}_{n\geq 0}$ *and the set of* $S^2 \times S^2$*-stable homotopy classes of lifts* $M\#^n S^2 \times S^2 \to BO$ *of the stable tangent microbundles of the same collection of manifolds.*

(2) Let $h\colon M \to N$ *be a homeomorphism of smooth 4-manifolds. Suppose that the diagram*

$$\begin{tikzcd} M \arrow[r, "\tau_M"] \arrow[d, "h"'] & BO \\ N \arrow[ur, "\tau_N"] & \end{tikzcd}$$

commutes up to homotopy. Let

$$G\colon M \times [0,1] \to BO$$

be a homotopy between τ_M *and* $\tau_N \circ h$*. Then there exists* $n \geq 0$ *so that*

$$h\#\,\mathrm{Id}\colon M\#^n S^2 \times S^2 \to N\#^n S^2 \times S^2$$

is concordant to a diffeomorphism. That is, there is a homeomorphism

$$H\colon (M\#^n S^2 \times S^2) \times [0,1] \to (N\#^n S^2 \times S^2) \times [0,1]$$

such that

$$H| = h\#\,\mathrm{Id}\colon (M\#^n S^2 \times S^2) \times \{0\} \to (N\#^n S^2 \times S^2) \times \{0\}$$

and such that

$$H|\colon (M\#^n S^2 \times S^2) \times \{1\} \to (N\#^n S^2 \times S^2) \times \{1\}$$

is a diffeomorphism. Moreover the homotopies

$$\tau_{(N\#^n S^2\times S^2)\times[0,1]} \circ H\colon (M\#^n S^2 \times S^2) \times [0,1] \to BO$$

and

$$G\#_{[0,1]}\,\mathrm{Id}\colon (M \times [0,1])\#_{[0,1]}((\#^n S^2 \times S^2) \times [0,1]) \to BO$$

are homotopic relative to $(M\#^n S^2 \times S^2) \times \{0,1\}$*, where* $\#_{[0,1]}$ *denotes connected sum along embedded intervals* $\{\mathrm{pt}\} \times [0,1]$ *and we have the natural identification*

$$(M \times [0,1])\#_{[0,1]}((\#^n S^2 \times S^2) \times [0,1]) = (M\#^n S^2 \times S^2) \times [0,1].$$

Moreover, if M *and* N *are compact and simply connected, then the homeomorphism* H *can be taken to be an isotopy i.e. to be level preserving.*

The proof applies the handle straightening theorem (Section 21.4.2) and the controlled h-cobordism theorem (Section 21.4.1). The final sentence of (2), upgrading concordance to isotopy, uses [**Qui86**].

21.4.6. $TOP(4)/O(4) \to TOP/O$ **is 5-connected.** Recall that the spaces $TOP(4)/O(4)$ and TOP/O are by definition the homotopy fibres in the fibration sequences:

$$\begin{tikzcd} TOP(4)/O(4) \arrow[r] \arrow[d] & BTOP(4) \arrow[r] \arrow[d] & BO(4) \arrow[d] \\ TOP/O \arrow[r] & BTOP \arrow[r] & BO. \end{tikzcd}$$

Theorem. *The stabilisation map* $TOP(4)/O(4) \to TOP/O$ *is 5-connected.*

The statement that $TOP(4)/O(4) \to TOP/O$ is 3-connected is a trivial consequence. However, 3-connectedness, which is enough for many upcoming applications, can be deduced more easily, whence the separate box for it in the flowchart (Figure 21.1). Quinn deduces it from the handle straightening theorem (Section 21.4.2), as described in [**Qui82b**, Theorem 2.2.3].

Part (1) of the sum stable smoothing theorem [**FQ90**, Theorem 8.6] from Section 21.4.5 is needed to prove that $TOP(4)/O(4) \to TOP/O$ is 4-connected, as in [**FQ90**, Theorem 8.7A]; see also [**LT84**] for 4-connectedness. The proof of [**FQ90**, Theorem 8.7A] uses part (2) of the sum stable smoothing theorem to prove 5-connectedness.

21.4.7. Smoothing away from a point.

Theorem (Noncompact 4-manifolds are smoothable). *Let M be a 4-manifold for which every connected component is noncompact. Then M admits a smooth structure. In particular, every closed, connected 4-manifold admits a smooth structure on the complement of a single point.*

Here is a sketch of the proof. First note that a noncompact 4-manifold is homotopy equivalent to a 3-dimensional CW complex as follows. By [**KS77**, Section III.4], such a manifold M can be embedded in $\mathbb{R}^N$ for some large N. By taking a normal disc bundle with a PL triangulation, it can be shown that M is homotopy equivalent to a finite dimensional CW complex. By [**Wal65a**, Theorem E], combined with noncompact Poincaré duality, M is homotopy equivalent to a 3-complex.

Now, by immersion theory (Section 21.2.1), a connected, noncompact 4-manifold admits a smooth structure if the classifying map $M \to BTOP(4)$ of the tangent microbundle lifts to a map $M \to BO(4)$, which occurs if and only if the composition

$$M \to BTOP(4) \to B(TOP(4)/O(4))$$

is null-homotopic. Here we use that the homotopy fibre $TOP(4)/O(4)$ is a *homotopy-everything H-space* [**BV68**, **BV73**], these days called an E_∞ *space*, and so admits a de-looping $B(TOP(4)/O(4))$ with $\Omega B(TOP(4)/O(4)) \simeq TOP(4)/O(4)$.

The composition

$$M \to B(TOP(4)/O(4)) \to B(TOP/O) \simeq K(\mathbb{Z}/2, 4)$$

is null-homotopic since M is homotopy equivalent to a CW complex, so we may identify $[M, B(TOP/O)] = H^4(M; \mathbb{Z}/2) = 0$. Write F for the homotopy fibre of the map $B(TOP(4)/O(4)) \to B(TOP/O)$. We have $\pi_i(F) \cong \pi_i(TOP/O, TOP(4)/O(4))$ for every i. Use that F is 3-connected and that M is homotopy equivalent to a 3-complex, to see that every map $M \to F$ is null-homotopic. It then follows from the exact sequence in sets

$$[M, F] \to [M, B(TOP(4)/O(4))] \to [M, B(TOP/O)]$$

that the map $M \to B(TOP(4)/O(4))$ we started with is null-homotopic. Therefore M admits a smooth structure, as claimed.

The handle straightening theorem provides an alternative route to the smoothing theorem, as in [**FQ90**, Theorem 8.2]. There is an analogue of this result for 5-dimensional cobordisms [**FQ90**, Corollary 8.7D]. The proof also uses that $TOP(4)/O(4) \to TOP/O$ is 3-connected.

Theorem (5-dimensional cobordisms are smoothable away from an arc). *Let $(W; M_0, M_1)$ be a 5-dimensional cobordism. There is a proper 1-dimensional submanifold $T \subseteq W$ with a normal bundle, such that $(W \smallsetminus T; M_0 \smallsetminus T, M_1 \smallsetminus T)$ admits a smooth structure.*

21.4.8. Normal bundles. In [**FQ90**, Section 9.3; **Qui82b**, Theorem 2.5.1], it is shown that normal bundles for locally flat submanifolds of 4-manifolds exist and are unique.

Let M be a 4-manifold and let S be a proper k-dimensional submanifold for some $k \geq 0$. A *normal bundle* for S is a pair $(E, p\colon E \to S)$ with the following properties.

(1) The total space E is a codimension zero submanifold of M.
(2) The map $p\colon E \to S$ is an $(4-k)$-dimensional vector bundle such that $p(x) = x$ for all $x \in S$.
(3) We have $\partial E = p^{-1}(\partial S)$.
(4) E is *extendable*, which means that given an $(n-k)$-dimensional vector bundle $(F, q\colon F \to S)$, and a radial homeomorphism of E to an open convex disc bundle in F, then the inclusion $E \hookrightarrow M$ can be extended to an embedding $F \to M$, satisfying $\partial F = q^{-1}(\partial S)$ and such that $E \to F \to M$ agrees with the inclusion $E \to M$.

THEOREM (Existence and uniqueness of normal bundles). *Every locally flat, proper submanifold S of a topological 4-manifold M has a normal bundle, unique up to ambient isotopy.*

The proof given in [**FQ90**, Section 9.3] uses the controlled h-cobordism theorem (Section 21.4.1), and that noncompact connected 4-manifolds can be smoothed (Section 21.4.7). Let S and M be as above, with S connected. Roughly, the proof of existence of normal bundles begins with taking the product of the ambient manifold M with the real line. The high dimensional analogue of the normal bundle theorem [**Qui79**, Theorem 3.4.1; **BS70**; **Cha79**; **Fer79**] provides a normal bundle of $S \times \mathbb{R}$ within $M \times \mathbb{R}$. A subset of this normal neighbourhood, with $S \times \mathbb{R}$ removed, is a controlled h-cobordism between the intersection with $M \times \{0\}$ and the restriction of the normal bundle of $S \times \mathbb{R}$ to $S \times \{1\}$. The latter is the complement of the zero section in a D^{4-k}-bundle on S, while the former lies in $M \times \{0\}$. The only missing ingredient now is a smooth structure, since we only have the controlled h-cobordism theorem with smooth input. This follows from the fact that noncompact, connected 4-manifolds are smoothable (Section 21.4.7). This implies that either of the two boundary components of the controlled h-cobordism is smooth, and since the h-cobordism deformation retracts to either boundary component, the smooth structure extends. To see this, let Q be a boundary component and let R be the h-cobordism, and note that the null homotopy of the map $Q \to B(TOP/O)$ coming from the fact that Q is smooth induces a null homotopy of the map $R \to B(TOP/O)$. Then high dimensional smoothing theory [**KS77**, Essay V] implies that R admits a smooth structure. A judicious choice of ε in the smooth input controlled h-cobordism theorem gives a product structure on the subset of the 5-dimensional normal neighbourhood that extends over $S \times \mathbb{R}$, and the intersection with $M \times \{0\} = M$ is the desired normal bundle of S in M.

21.4.9. Topological transversality and map transversality. Quinn's work was the final step in establishing topological transversality in all dimensions and codimensions [**Qui88**]. The proof in the 4-dimensional case appears in [**FQ90**, Section 9.5], combining the handle straightening theorem (Section 21.4.2), the existence of normal bundles (Section 21.4.8), the submanifold smoothing theorem [**FQ90**, Theorem 8.7C], and the disc embedding theorem with smooth input (Section 21.1). The proof of the submanifold smoothing theorem [**FQ90**, Theorem 8.7C] uses that the stabilisation map $TOP(4)/O(4) \to TOP/O$ is 3-connected (Section 21.4.6)

THEOREM (Topological transversality). *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.*

Above, *transverse* means that the points of intersection have coordinate neighbourhoods within which the submanifolds appear as transverse linear subspaces. The formulation above is sometimes called *submanifold transversality*.

Map transversality for all dimensions and codimensions, stating that a map can be perturbed so that the inverse image of a submanifold is a submanifold of the same codimension, can be deduced from submanifold transversality for all dimensions and codimensions, using a general argument not specific to dimension four. See [**FNOP19**, Section 10] for details of this deduction. In high dimensions, a submanifold need not have a normal microbundle let alone a normal bundle. To state map transversality we need to at least use the notion of a normal microbundle. The statement of the theorem and the next definition are from [**FNOP19**].

Let $f\colon M \to N$ be a continuous map and let X be a submanifold of N with normal microbundle νX. The map f is said to be *transverse* to νX if $f^{-1}(X)$ is a submanifold admitting a normal microbundle $\nu f^{-1}(X)$ with data $f^{-1}(X) \xrightarrow{i} \nu f^{-1}(X) \xrightarrow{r} f^{-1}(X)$ (see Definition 21.1), so that

$$\begin{aligned} f\colon \nu f^{-1}(X) &\to f^*\nu X \\ m &\mapsto (r(m), f(m)) \end{aligned}$$

is an isomorphism of microbundles.

In the following theorem there are no restrictions on dimensions nor codimensions, and M and N are topological manifolds.

THEOREM (Map transversality in all dimensions). *Let $Y \subseteq N$ be a locally flat, proper submanifold with normal microbundle νY. Let $f\colon M \to N$ be a map and let U be a neighbourhood of the set*

$$\operatorname{graph} f \cap (M \times Y) \subseteq M \times N.$$

Then there exists a homotopy $F\colon M \times [0,1] \to N$ such that

(1) $F(m,0) = f(m)$ for all $m \in M$;
(2) $F_1\colon M \to N$ taking each $m \mapsto F(m,1)$ is transverse to νY; and
(3) for $m \in M$ either
  (i) $(m, f(m)) \notin U$, in which case $F(m,t) = f(m)$ for all $t \in [0,1]$, or
  (ii) $(m, f(m)) \in U$, in which case $(m, F(m,t)) \in U$ for all $t \in [0,1]$.

In other words, the function f is homotopic to a function F_1 so that $F_1^{-1}(Y)$ is a submanifold of M with codimension $\dim(N) - \dim(Y)$.

The new cases proved using the 5-dimensional controlled h-cobordism theorem (Section 21.4.1) are (i) when N is 4-dimensional; this uses the transversality for submanifolds stated above, and (ii) when Y is 4-dimensional; this uses the submanifold smoothing theorem [**FQ90**, Theorem 8.7C], which itself relies on the statement that the stabilisation map $TOP(4)/O(4) \to TOP/O$ is 3-connected (Section 21.4.6).

21.4.10. The immersion lemma. We recall first the definition of a framed immersion from Chapter 11.

A *framed immersion* of an orientable surface F in a smooth 4-manifold M is an immersion of F in M such that the normal bundle of the image of F is trivial. Framed immersions can also be defined topologically as follows [**FQ90**, Section 1.2]. Given an abstract surface F, form the product $F \times \mathbb{R}^2$. Consider disjoint copies D and E of $\mathbb{R}^2$ in F. Perform a plumbing operation on $D \times \mathbb{R}^2$ and $E \times \mathbb{R}^2$ to obtain a plumbed model. A (topological) *framed immersion* of the abstract surface F in a topological 4-manifold M is a map from a plumbed model for F to some open set in

M, that is a homeomorphism onto its image. Such a homeomorphism determines a map $g\colon F \to M$ when we restrict to the image of F in the framed model and we say that the map g *extends to a framed immersion*.

If the boundary of F is nonempty, then we will usually have a fixed framing already prescribed on the boundary. In this case, we say that a map $g\colon F \to M$ *extends to a framed immersion* if it extends to a framed immersion restricting to the given framing on ∂F.

The following theorem appears as [**FQ90**, Lemma 1.2(1)].

THEOREM (Immersion lemma). *Let $f\colon F \to M$ be a continuous map from a compact, orientable surface F to a 4-manifold M. Suppose that f extends to a framed immersion in a neighbourhood of ∂F. After changing the given framing on ∂F by rotations, there is a homotopy of f fixing ∂F to a map $F \to M$ that extends to a framed immersion.*

Above, for a space N, if $r\colon N \to O(2)$ is a map, we can define an automorphism of $N \times D^2$ by acting by $r(n)$ on $\{n\} \times D^2$. Two framings *differ by rotations* if one is obtained from the other by composing with such an automorphism on $N \times D^2$.

One can decompose a nonorientable surface into a union of orientable surfaces, such as discs and bands, and so apply this theorem to obtain topological immersions of nonorientable surfaces as well.

The proof of the immersion lemma uses that 4-manifolds can be smoothed away from a point (Section 21.4.7) – once we have a smooth structure, we are able to use standard facts from differential topology to finish the proof (see [**FQ90**, Corollary 9.5C]).

21.4.11. Handle decompositions of 5-manifolds. Let W be a 5-manifold with boundary and let $N \subseteq \partial W$ be a codimension zero submanifold of ∂W. Either or both of N and ∂W may be empty. A *relative handle decomposition* of (W, N) is a (possibly infinite) filtration

$$W_0 \subseteq W_1 \subseteq W_2 \subseteq \cdots$$

of W such that

(1) W_0 is a collar $N \times [0,1]$.
(2) W_{i+1} is obtained from W_i by adding a handle to $\partial W_i \smallsetminus N \times \{0\}$, meaning that for some $k \in \{0, \dots, 5\}$ there is a topological embedding
$$\phi\colon S^{k-1} \times D^{5-k} \to \partial W_i \smallsetminus N \times \{0\}$$
and a pushout:
$$\begin{array}{ccc} S^{k-1} \times D^{5-k} & \xrightarrow{\phi} & W_i \\ \downarrow & & \downarrow \\ D^k \times D^{5-k} & \longrightarrow & W_{i+1} \end{array}$$
(3) Every $x \in W$ has a neighbourhood U that intersects only finitely many of the sets $W_{i+1} \smallsetminus W_i$.

Compare with Section 13.2.

The existence of these structures on any 5-manifold was also proven by Quinn in [**Qui82b**]. Note that 4-manifolds with a handle decomposition have a smooth structure, since handles are attached along 3-manifolds and so the attaching maps can be smoothed. Thus there are 4-manifolds, such as the E_8-manifold, which admit no handle decomposition. Indeed, an n-manifold M has a topological handle decomposition if and only if M is not a non-smoothable 4-manifold [**KS77**, Essay 3, Section 2; **Bin59**, Theorem 8; **Moi52a**].

THEOREM (5-manifolds have handle decompositions). *Let (W, N) be a 5-dimensional topological manifold pair, with $N \subseteq \partial W$. Then W admits a topological handle decomposition relative to N.*

The theorem is proven in [**Qui82b**, Theorem 2.3.1; **FQ90**, Section 9.1]. The proof uses that 5-dimensional cobordisms have smooth structures away from a union of arcs (Section 21.4.7), which uses the fact that $TOP(4)/O(4) \to TOP/O$ is 3-connected, and the controlled h-cobordism theorem (Section 21.4.1).

21.5. Category preserving theorems

The category preserving results in this development are shown in red in Figure 21.1. We outline their proofs.

As previously mentioned, to prove the fully topological disc embedding theorem, where the ambient manifold is not required to be smooth, one repeats the proof described in this book but in the purely topological setting. One must then contemplate precisely where smoothness appears in the argument.

The immersion lemma (Section 21.4.10) is needed to approximate continuous functions by generic immersions, such as when we assert the existence of immersed Whitney discs due to the vanishing of intersection numbers, or more generally, when we assume that the track of a null homotopy yields an immersed disc.

In order to take parallel copies, for instance when tubing into a geometrically transverse sphere to remove intersections, we need the existence of normal bundles (Section 21.4.8) for generically immersed surfaces in a topological 4-manifold. We also need such normal bundles in the very first step of the proof, where we pass to the complement of a neighbourhood of the given immersed discs (after a regular homotopy) and apply Proposition 16.2.

We need transversality (Section 21.4.9) in the topological category when we produce a new Whitney disc or cap and perturb it so that it intersects all previous constructions transversely. Without transversality, if one attempted to apply the immersion lemma directly, one would also need to perturb the previous constructions, thereby potentially disrupting properties already arranged, for example that surface stages of a grope are embedded. These are the only places we used the smoothness of the ambient 4-manifold in the proof of the disc embedding theorem with smooth input.

The analogues of the immersion lemma, the existence of normal bundles, and transversality are all standard tools in the smooth category. They can all be derived in the topological category from [**Qui82b**], as described in the previous section. Their topological versions allow one to upgrade the category losing results proved in this book to fully topological results.

THEOREM (Disc embedding theorem, topological input). *Let M be a connected, topological 4-manifold, with nonempty boundary and with $\pi_1(M)$ a good group. Let $f_1, \dots, f_n \colon (D^2, S^1) \looparrowright (M, \partial M)$ be a properly immersed collection of discs in M with pairwise disjoint, embedded boundaries. Suppose there is an immersed collection $g_1, \dots, g_n \colon S^2 \looparrowright M$ of framed dual spheres such that $\lambda(f_i, g_j) = \delta_{ij}$ and $\lambda(g_i, g_j) = 0 = \mu(g_i)$ for all $i, j = 1, \dots, n$. Then there exist pairwise disjoint, flat embedded discs $\overline{f}_1, \dots, \overline{f}_n \colon D^2 \hookrightarrow M$, with an immersed collection of framed geometrically transverse spheres $\overline{g}_1, \dots, \overline{g}_n \colon S^2 \looparrowright M$, such that for every i the discs $\overline{f}_i$ and f_i have the same framed boundary and the sphere $\overline{g}_i$ is homotopic to g_i.*

Next, consider the proof of the compact s-cobordism theorem with smooth input (Theorem 20.1). The smoothness of the given cobordism provides a handle decomposition. In the topological category, the existence of (topological) handle decompositions for 5-manifolds was shown by Quinn [**Qui82b**] (Section 21.4.11).

Thereafter, the proof uses once again the three fundamental tools highlighted in blue in the flowchart (Figure 21.1), as well as the category preserving disc embedding theorem.

THEOREM (Compact s-cobordism theorem, topological input). *Let N be a compact, topological, 5-dimensional h-cobordism between closed 4-manifolds M_0 and M_1 with vanishing Whitehead torsion $\tau(N, M_0)$. Suppose that $\pi_1(N)$ is a good group. Then N is homeomorphic to the product $M_0 \times [0,1]$. In particular, M_0 and M_1 are homeomorphic.*

The category preserving disc embedding theorem and the three fundamental tools are the only necessary ingredients needed to repeat the proof of the sphere embedding theorem (Theorem 20.4) in the topological category.

THEOREM (Sphere embedding theorem, topological input). *Let M be a connected, topological 4-manifold such that $\pi_1(M)$ is good. Suppose there exists an immersed collection $f_1, \dots, f_n \colon S^2 \looparrowright M$, with $\lambda(f_i, f_j) = 0$ for every $i \neq j$ and $\mu(f_i) = 0$ for $i = 1, \dots, n$. Suppose moreover that there is an immersed collection $g_1, \dots, g_n \colon S^2 \looparrowright M$ of framed dual spheres with $\lambda(f_i, g_j) = \delta_{ij}$ for all $i, j = 1, \dots, n$. Then there exist pairwise disjoint, locally flat, embedded spheres $\overline{f}_1, \dots, \overline{f}_n \colon S^2 \hookrightarrow M$, with each $\overline{f}_i$ regularly homotopic to f_i, and with an immersed collection of framed geometrically transverse spheres $\overline{g}_1, \dots, \overline{g}_n \colon S^2 \looparrowright M$, with each $\overline{g}_i$ homotopic to g_i.*

21.5.1. The surgery sequence for good groups. The primary application of the fully topological sphere embedding theorem is to surgery theory. In this section, we briefly state the key definitions and the main result. We will discuss the surgery sequence in detail in Chapter 22, where we will explain the notation, the terms in the sequence, the maps, and why the sequence is exact.

Elements of the *structure set* $\mathcal{S}^s(X)$ of a 4-dimensional *Poincaré complex* X are represented by pairs (M, f), where M is a closed topological 4-manifold and $f \colon M \to X$ is a simple homotopy equivalence, up to s-cobordism over $X \times [0,1]$. The *normal maps* $\mathcal{N}(X)$ are represented by quadruples (M, f, ξ, b) where $f \colon M \to X$ is a *degree one normal map* from a closed, topological 4-manifold M to X mapping a fundamental class $[M]$ to the fundamental class $[X]$, together with a stable vector bundle $\xi \to X$ and a bundle map $b \colon \nu_M \to \xi$, considered up to normal bordism over $X \times [0,1]$. The *L-groups* $L_4^s(\mathbb{Z}[\pi_1(X)])$ and $L_5^s(\mathbb{Z}[\pi_1(X)])$ are algebraically defined, and only depend on $\pi_1(X)$ and the orientation character $w \colon \pi_1(X) \to \{\pm 1\}$. The group $L_4^s(\mathbb{Z}[\pi_1(X)])$ consists of stable isomorphism classes of sesquilinear, hermitian, nonsingular simple forms, while $L_5^s(\mathbb{Z}[\pi_1(X)])$ consists of equivalence classes of nonsingular quadratic formations.

THEOREM (The surgery sequence is defined and exact). *For X a 4-dimensional Poincaré complex with $\pi_1(X)$ good, the structure set $\mathcal{S}^s(X)$ is nonempty if and only if the Spivak normal fibration of X admits a topological vector bundle reduction for which the associated degree one normal map has vanishing surgery obstruction in $L_4^s(\mathbb{Z}[\pi_1(X)])$. If $\mathcal{S}^s(X)$ is nonempty, the action of $L_5^s(\mathbb{Z}[\pi_1(X)])$ on $\mathcal{S}^s(X)$ is defined, and the surgery sequence*

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

is an exact sequence of pointed sets.

The proof uses the sphere embedding theorem with topological input to show both the exactness at $\mathcal{N}(X)$ and to define the action of $L_5^s(\mathbb{Z}[\pi_1(X)])$ on $\mathcal{S}^s(X)$. The proof of exactness at the structure set uses the existence of handle decompositions on topological 5-manifolds (Section 21.4.11).

The computation of the *normal invariants*, that is identifying the set $\mathcal{N}(X)$, is often critical in applications. We will explain the notation further and give an example of this in Chapter 22. For now, let $G(k)$ be the monoid of self-homotopy equivalences of S^{k-1}, and let G be the colimit of the system with $G(k) \to G(k+1)$ induced by suspension. Let $BG(k)$ and BG be the associated classifying spaces. The space G/TOP is by definition the homotopy fibre in the fibration sequence $G/TOP \to BTOP \to BG$.

THEOREM ($\mathcal{N}(X) = [X, G/TOP]$). *For a 4-dimensional Poincaré complex X, there are bijections*

$$\mathcal{N}(X) \cong [X, G/TOP] \cong H^2(X; \mathbb{Z}/2) \oplus H^4(X; \mathbb{Z}).$$

The identification of normal bordism classes of degree one normal maps over X with $[X, G/TOP]$ uses the Pontryagin-Thom construction [**Lüc02**, Chapter 3.3]. This uses map transversality, in particular surjectivity of $\mathcal{N}(X) \to [X, G/TOP]$ uses the case of map transversality where the inverse image has dimension four.

There is a 5-connected map $G/TOP \to K(\mathbb{Z}/2, 2) \times K(\mathbb{Z}, 4)$ [**KT01**, p. 397], from which we obtain the identification $[X, G/TOP] \cong H^2(X; \mathbb{Z}/2) \oplus H^4(X, \mathbb{Z})$. See Section 22.3.2 for further details.

21.6. Flagship results

Some landmark results in topological 4-manifold theory are shown in yellow in Figure 21.1.

21.6.1. Classification of closed, simply connected 4-manifolds. The classification of closed, topological, simply connected 4-manifolds, up to homeomorphism, realises the "dream", quoting [**Fre82a**], "that some key principle from the high dimensional theory would extend . . . to dimension four, and bring with it the beautiful adherence of topology to algebra familiar in dimensions greater than or equal to five."

This is in stark contrast to the behaviour of smooth 4-manifolds which, as demonstrated by Donaldson, do not adhere so closely to algebra.

THEOREM (Classification of closed, topological, simply connected 4-manifolds). *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, 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 (in other words, their Kirby-Siebenmann invariants coincide). Then there is a homeomorphism $G\colon M \to M'$ such that $G_* = \phi\colon H_2(M; \mathbb{Z}) \to H_2(M'; \mathbb{Z})$.

Moreover, according to [**Qui86**], the homeomorphism G is uniquely determined up to isotopy by ϕ.

A version of this was proven in [**Fre82a**, Theorem 1.5], restricted to closed, topological, simply connected 4-manifolds that are smoothable away from a point. Quinn proved that any connected, noncompact 4-manifold is smoothable (Section 21.4.7), thereby showing that Freedman's original theorem applies to all closed, topological, simply connected 4-manifolds. Freedman's proof of uniqueness used the proper h-cobordism theorem with smooth input and finitely many ends (in fact the case of a single end). As explained previously, virtually all of Freedman's work in [**Fre82a**] required a smooth ambient 4-manifold, whence the (apparent) restriction to 4-manifolds smoothable away from a point in [**Fre82a**].

Begin with closed, topological, simply connected 4-manifolds M and M' with the same intersection form, as well as the same Kirby-Siebenmann invariant in case the intersection form is odd, such that $M \smallsetminus \{p\}$ and $M' \smallsetminus \{p'\}$ are smoothable, for some points $p \in M$ and $p' \in M'$. By the Milnor-Whitehead classification [**Whi49**, **Mil58**], M and M' are homotopy equivalent. Since they additionally have the same Kirby-Siebenmann invariant, there exists a topological h-cobordism W between M and M'. Constructing this cobordism uses that the map $\mathcal{N}(X) \to [X, G/TOP]$ is injective, a fact that only uses high dimensional topological transversality [**KS77**] and is thus independent from the work of Quinn. Surjectivity of this map, as discussed in Section 21.5.1, uses 4-dimensional map transversality.

Remove a properly embedded arc from W joining p and p'. The result is a proper h-cobordism which is simply connected at infinity and which can be smoothed using immersion theory. Then apply the proper h-cobordism theorem proved by Freedman [**Fre82a**, Theorem 10.3] (see Section 21.3.1) to conclude that $M \smallsetminus \{p\}$ is homeomorphic to $M' \smallsetminus \{p'\}$. Extend the homeomorphism over one point compactifications to conclude that M is homeomorphic to M'.

Freedman's proof of existence of closed, topological, simply connected 4-manifolds realising a given intersection form applies the same strategy used to construct the E_8-manifold in Section 21.3.2, with the key ingredient being that every integral homology 3-sphere bounds some contractible 4-manifold. The proof applies directly in the case of any even intersection form. In the case of odd intersection forms, one must construct both a stably smoothable and a non-stably smoothable representative. The case of the intersection form $[+1]$ was done in Chapter 1.

Given an odd candidate intersection form θ, construct a framed link L in S^3 whose linking-framing matrix represents θ. Let Σ be the homology 3-sphere obtained from Dehn surgery on S^3 along L using the given framings. Let $J \subseteq L$ be a *characteristic sublink*, meaning that, for every component L_i of L, the total linking number $\ell\mathrm{k}(J, L_i) = \sum_k \ell\mathrm{k}(J_k, L_i) \mod 2$ equals the framing of L_i, where the $\{J_k\}$ are the components of L. The set of characteristic sublinks of a given surgery diagram are in one to one correspondence with the spin structures on the corresponding 3-manifold. In this case, there is a unique spin structure since we have an integral homology sphere, so there is a unique characteristic sublink. See, for example, [**GS99**, Propositions 5.6.3 and 5.7.11] for the last two assertions. Let L' be a link obtained from L by tying in a local 0-framed knot with nontrivial Arf invariant into a component of J.

Construct closed manifolds M and M' by first attaching 2-handles to D^4 along the components of L and L' in S^3 respectively, as dictated by the linking-framing matrix. Then cap off the resulting homology sphere boundaries Σ and Σ' by the contractible, topological 4-manifolds C and C' provided by Proposition 22.6 (see also Section 21.3.2). It follows from [**FK78**] (see also [**CST12**, Theorem 2]) that the Kirby-Siebenmann invariants of M and M' differ. Indeed the formula in [**CST12**, Theorem 2] computes the Kirby-Siebenmann invariants of M and M' explicitly. Since L and L' have the same linking-framing matrix, the associated 4-manifolds

M and M' have isometric intersection forms. The manifold with vanishing Kirby-Siebenmann invariant is stably smoothable, and the other is not.

An alternative proof of the classification theorem using the surgery sequence is described in detail in Theorem 22.3. Apart from the surgery sequence, this proof requires the existence of the E_8-manifold, the category preserving s-cobordism theorem, and the computation of the normal invariants. In particular, note that the existence of the E_8-manifold is needed prior to proving the classification theorem.

As mentioned in the introduction, further classification results have been obtained, such as for compact, topological, simply connected 4-manifolds with fixed connected boundaries [**Boy86**, **Boy93**, **Vog82**, **Sto93**] as well as for closed, topological 4-manifolds with infinite cyclic fundamental group [**FQ90**, Theorem 10.7A, page 173]. See Section 22.3.4 for further discussion and references.

21.6.2. The Poincaré conjecture with topological input.

THEOREM (Poincaré conjecture). *Let Σ be a closed, topological 4-manifold homotopy equivalent to S^4. Then Σ is homeomorphic to S^4.*

This category preserving Poincaré conjecture was first proved in [**Fre82a**, Theorem 1.6] as a direct corollary of the classification result of the previous section. Since a topological homotopy 4-sphere is smoothable away from a point by immersion theory, this is independent of the work of Quinn.

We now discuss three further routes, which also bypass the full classification.

PROOF OF THE POINCARÉ CONJECTURE FOLLOWING [**Fre82a**]. Given a topological homotopy 4-sphere Σ, the punctured manifold $\Sigma \smallsetminus \{\mathrm{pt}\}$ has empty boundary and is proper homotopy equivalent to $\mathbb{R}^4$. Thus by Section 21.3.1, $\Sigma \smallsetminus \{\mathrm{pt}\}$ is homeomorphic to $\mathbb{R}^4$. Extend the homeomorphism over the one-point compactifications to conclude that Σ is homeomorphic to S^4.

We remark that this proof only uses results from [**Fre82a**], but bypasses the classification result proved there. □

Let $k \geq -1$. A subset $A \subseteq X$ of a space X, not necessarily a manifold, is said to be *k-LC embedded*, that is, locally k-connected and embedded (sometimes called *k-LCC*, that is, locally k-coconnected) if for all $a \in A$, and every neighbourhood U of a in X, there is a neighbourhood $V \subseteq U$, such that every map $S^k \to V \smallsetminus A$ extends to a map $D^{k+1} \to U \smallsetminus A$. In the case $k = 1$, this means that every loop in $V \smallsetminus A$ is null-homotopic in $U \smallsetminus A$.

The subset A is said to be *locally homotopically unknotted* if for all $a \in A$, and every neighbourhood U of a in X, there is a neighbourhood $V \subseteq U$ such that the inclusion induced map $\pi_1(V \smallsetminus A) \to \pi_1(U \smallsetminus A)$ has abelian image and A is k-LCC for every $k \neq 1$. See [**DV09**, Section 1.3] for further generalisations and applications of this notion.

PROOF OF THE POINCARÉ CONJECTURE FOLLOWING [**FQ90**]. This argument appears in [**FQ90**, Corollary 7.1B] and uses the fully topological s-cobordism theorem. Given a topological homotopy sphere Σ, consider the cone $\mathscr{C}\Sigma = \Sigma \times [0,1]/\Sigma \times \{1\}$. The cone point p, given by the image of $\Sigma \times \{1\}$, is the only potential non-manifold point. However, it is 1-LC embedded, since it has arbitrarily small neighbourhoods U with $U \smallsetminus \{p\}$ homotopy equivalent to Σ. By [**BL78**, Theorem 1.5], the cone point is thus a manifold point. Remove a small ball neighbourhood of the cone point from $\mathscr{C}\Sigma$ to produce a simply connected, topological h-cobordism between Σ and S^4. By the category preserving s-cobordism theorem, the cobordism is a product and so Σ is homeomorphic to S^4. □

Proof of the Poincaré conjecture. This final proof is compatible with the one we gave for the Poincaré conjecture with smooth input (Theorem 20.3).

Begin with a closed topological homotopy 4-sphere Σ. The 4-manifold Σ is spin, since $H^2(\Sigma;\mathbb{Z}/2)=0$. The Kirby-Siebenmann invariant for a closed spin 4-manifold equals the signature divided by 8 (see, for example, [**FNOP19**, Theorem 8.2(4)]), so $\mathrm{ks}(\Sigma)=0$. By the sum stable smoothing theorem (see Section 21.4.5), there is an $n\geq 0$ such that $\Sigma\#^n(S^2\times S^2)$ admits a smooth structure. The manifolds Σ and $\Sigma\#^n(S^2\times S^2)$ cobound a topological, stably framed cobordism built from $\Sigma\times[0,1]$ and $\natural^n S^2\times D^3$. The signature of Σ vanishes since $H_2(\Sigma;\mathbb{Z})=0$. Thus, the smooth, spin 4-manifold $\Sigma\# n(S^2\times S^2)$ has vanishing signature and therefore bounds a compact, stably framed 5-manifold. As a result, the homotopy sphere Σ bounds a compact, topological, stably framed 5-manifold W. The rest of the proof of Theorem 20.3 goes through verbatim, except that we apply the fully topological s-cobordism theorem at the end. □

21.6.3. Alexander polynomial one knots are slice. The Alexander polynomial of a knot in S^3 is a Laurent polynomial in $\mathbb{Z}[t,t^{-1}]$, denoted $\Delta_K(t)$. It is well defined up to multiplication by the units $\pm t^k$. We write $\doteq$ for the relation where two Laurent polynomials are equivalent if and only if they are related by multiplication with a unit. The following is a famous application of Freedman's work in the realm of knot theory.

Theorem. *Let $K\subseteq S^3$ be a knot with Alexander polynomial $\Delta_K(t)\doteq 1$. Then K is topologically slice, that is K is the boundary of a locally flat disc properly embedded in D^4.*

A proof using surgery theory appeared in [**Fre84**, Theorem 7; **FQ90**, Theorem 11.7B] and is described in Chapter 1 (Theorem 1.14). The argument uses the existence of the E_8-manifold, the topological surgery sequence in the case of infinite cyclic fundamental group, as well as the topological input Poincaré conjecture.

Remark 21.4. The results [**Fre82a**, Theorems 1.13 and 1.14] are commonly, but erroneously, cited for the theorem that Alexander polynomial one knots are topologically slice. Indeed, the full statement above was never claimed by Freedman in [**Fre82a**]. Moreover, the results claimed rely on [**Fre82b**, Lemma 2], and a counterexample to this lemma was presented in [**GT04**].

Even ignoring the problem associated with [**Fre82b**, Lemma 2], [**Fre82a**] would not be the correct citation. The first theorem cited above, [**Fre82a**, Theorem 1.13], states that every Alexander polynomial one knot bounds an embedded disc in D^4. But the disc is only claimed to be locally homotopically unknotted. (Later work of Quinn [**FQ90**, Theorem 9.3A] shows that locally homotopically unknotted discs are indeed locally flat.) The other result cited, [**Fre82a**, Theorem 1.14], only asserts that the untwisted Whitehead double of a knot with Alexander polynomial one is topologically slice.

Since the first proof that Alexander polynomial one knots are slice makes crucial use of Quinn's developments, by working solely in the topological category, we suggest that this result ought to be attributed to both Freedman and Quinn.

Proof using the disc embedding theorem. The most direct route to slicing Alexander polynomial one knots is given in [**GT04**], using a single application of the disc embedding theorem with smooth input, where the ambient manifold has fundamental group $\mathbb{Z}$. The strategy is as follows. Begin with a Seifert surface Σ for a knot K with trivial Alexander polynomial, with its interior pushed into D^4. The complement of Σ has fundamental group $\mathbb{Z}$. Our goal is to ambiently surger Σ to produce a slice disc for K. Using the Alexander polynomial one condition, find a half basis of curves on Σ, bounding immersed discs $\{\Delta_i\}$ with interiors in the complement of Σ. The discs $\{\Delta_i\}$ are further equipped with algebraically transverse spheres produced using a dual basis of curves on Σ. Apply the disc embedding theorem with smooth input to the complement of Σ in D^4 to replace $\{\Delta_i\}$ by mutually disjoint, flat embedded discs in the complement of Σ, with the same framed boundary as the $\{\Delta_i\}$. Surger Σ using these embedded discs to produce a locally flat disc in D^4 bounded by K. This completes the proof. □

A third proof is possible using the geometric plus construction, as suggested at the bottom of [**FQ90**, p. 211], but this has not yet appeared in print.

For a knot K with Alexander polynomial one, the proofs above yield a locally flat slice disc Δ in D^4 such that $\pi_1(D^4 \setminus \Delta) \cong \mathbb{Z}$. In [**CP19**] it was shown that such discs are unique up to topological ambient isotopy relative to the boundary.

21.6.4. Slice knots. Donaldson's restrictions on the intersection forms of closed, smooth 4-manifolds (see Section 21.2.2), combined with the result that Alexander polynomial one knots are topologically slice, implies that there exist knots that are topologically slice but not smoothly slice. Such a knot bounds a locally flat, embedded disc in D^4, but no smoothly embedded disc therein. According to folklore, this was first observed, independently, by Casson and Akbulut, but neither published. See instead [**Gom86**, **CG88**].

We describe how Donaldson theory can be applied to obstruct smooth sliceness of knots, using the strategy of Casson, Akbulut, and [**CG88**]. If a knot K is slice, then $\Sigma_2(K)$, the double cover of S^3 branched along K, bounds a smooth rational homology ball. If $\Sigma_2(K)$ also bounds a smooth 4-manifold X with definite non-diagonalisable intersection form, such as $E_8 \oplus [+1]$, then the putative rational homology ball can be used to cap it off and produce a closed, smooth, oriented 4-manifold with nonstandard intersection form, thereby contradicting Donaldson's Theorem A. This implies that K is not smoothly slice. In the argument above, one may also replace $\Sigma_2(K)$ by the ± 1-framed surgery on S^3 along K, since if K were slice, this surgery would bound a contractible 4-manifold.

Applying this strategy to the case of a knot with trivial Alexander polynomial produces a knot that is topologically but not smoothly slice. For example, [**CG88**, Theorem 2.17] showed that for K the positive untwisted Whitehead double of the right handed trefoil, $\Sigma_2(K)$ bounds a smooth 4-manifold with definite, non-diagonalisable intersection form. It is easy to compute that the Alexander polynomial of every untwisted Whitehead double is one.

Remark 21.5. The reader may wonder why Rochlin's theorem cannot be used in the above strategy. Recall that Rochlin's theorem states that the signature of a closed, smooth, spin 4-manifold is divisible by 16. Let K be a topologically slice knot. The Rochlin invariant of the double branched cover of S^3 along K is trivial (see, for example, [**Sav02**, Theorem 2.17]), as is the Rochlin invariant of the ± 1-framed Dehn surgery on S^3 along along K (see, for example, [**Sav02**, Theorem 2.10; **GA70**]). This implies that any smooth, spin 4-manifold bounded by these 3-manifolds has signature divisible by 16, and at least the above strategy cannot be used to obstruct the smooth sliceness of K.

Donaldson's theorem is employed differently in [**Gom84**] and [**Gom86**], inspired by work of Kuga [**Kug84**], who used Donaldson's Theorem A to obstruct homology classes in $S^2 \times S^2$ from being represented by smoothly embedded 2-spheres.

Advances in gauge-theoretic methods, especially the availability of computable invariants from Heegaard-Floer homology, have since made the detection of topologically slice knots that are not smoothly slice relatively straightforward, and much subtle behaviour in the set of topologically slice knots modulo smoothly slice knots has been detected, such as in [**End95**, **OS03**, **Lob09**, **Ras10**, **HK12**, **CHH13**, **Hom14**, **HW16**, **HKL16**, **LL16**, **DV16**, **OSS17**, **CK17**].

21.6.5. Exotic $\mathbb{R}^4$s. Perhaps the most dramatic consequence of combining the work of Freedman and Donaldson is the existence of exotic versions of $\mathbb{R}^4$.

THEOREM (Existence of exotic $\mathbb{R}^4$s). *There exists a smooth 4-manifold $\mathcal{R}$ which is homeomorphic to $\mathbb{R}^4$ but not diffeomorphic to $\mathbb{R}^4$. Equivalently, there exists an exotic smooth structure on $\mathbb{R}^4$.*

This is especially striking since $\mathbb{R}^n$ has a unique smooth structure for all $n \neq 4$ [**Sta62**, Corollary 5.2; **Moi77**] (see also [**Sco05**, Chapter 4.5]).

The two primary sources of exotic $\mathbb{R}^4$s are the smooth failure of the h-cobordism theorem and the smooth failure of surgery. The proofs in the literature used Casson handles. We give details of the proofs below since we have modified them to use skyscrapers.

PROOF USING THE SMOOTH FAILURE OF SURGERY. Let M denote the smooth manifold $\mathbb{CP}^2\#9\overline{\mathbb{CP}}^2$. Let $\{h, e_1, \dots, e_9\}$ denote the canonical generators of $H_2(M;\mathbb{Z})$. It follows from elementary computation that the intersection form on $H_2(M;\mathbb{Z})$ is isomorphic to $-E_8 \oplus [-1] \oplus [+1]$. Alternatively, use the classification of indefinite, integral, bilinear forms [**Ser70**, **MH73**]. The class $\alpha := 3h + \sum_{i=1}^{8} e_i$ has square $+1$ and the orthogonal complement of $\{\alpha\}$ in $H_2(M)$ has definite intersection form $-E_8 \oplus [-1]$. The latter form is not diagonalisable over $\mathbb{Z}$. The class $\beta := -e_1 + e_9$ has square $+2$ and $\lambda(\alpha, \beta) = 1$.

The source, roughly speaking, of the exotic $\mathbb{R}^4$ will be that the class α can be represented by a topological embedding of a sphere, but cannot be represented by such a smooth embedding. If it could be, then blowing it down would yield a 4-manifold with intersection form $-E_8 \oplus [-1]$, which violates Donaldson's Theorem A.

Since M is simply connected and smooth, the classes α and β may be represented by immersed 2-spheres, A and B respectively. Apply the geometric Casson lemma (Lemma 15.3) to ensure A and B are geometrically transverse, at the expense of finger moves. Add local cusps to A and B to ensure that the Euler numbers of the normal bundles are $+1$ and $+2$ respectively. It follows that $\mu(A) = \mu(B) = 0$ using Proposition 11.8. Thus all the self-intersections of A may be paired by framed, immersed Whitney discs $\{W_i\}$ in M. Remove any intersections of $\{W_i\}$ with A by tubing into B. This changes the framing of the Whitney discs by multiples of $+2$, but this can be corrected by adding local cusps to the $\{W_i\}$.

Let $p \in A$ be a point away from all Whitney arcs and the intersection of A with B. Let $\mathring{M}$ denote the complement in M of a small open ball D around p. Assume D does not intersect B nor the Whitney discs. Let $\mathring{A}$ denote the complement $A \smallsetminus D$. Then we have a family of framed, immersed Whitney discs, which we still call $\{W_i\}$, pairing all the self-intersections of $\mathring{A}$, whose interiors lie in $N := \mathring{M} \smallsetminus A$.

Use Clifford tori at the double points of A to create transverse spheres for $\{W_i\}$ lying in N as usual. This involves surgering the Clifford tori using meridional discs tubed into B. We obtain framed spheres since each disc is used algebraically zero times. Now the discs $\{W_i\}$, equipped with these transverse spheres, satisfy the hypotheses of the disc embedding theorem. In particular, the group $\pi_1(N)$ is trivial, and hence good, because A is π_1-negligible.

However, we do not directly apply the disc embedding theorem. Rather, apply Proposition 17.12 to replace the discs $\{W_i\}$ by another family of discs $\{W_i'\}$, with vanishing intersection and self-intersection numbers, whose intersection and self-intersection points are paired by a union-of-discs-like 1-storey capped tower $\mathcal{T}$ with at least four surface stages. Let C denote the union

$$C := \nu\mathring{A} \cup \bigcup_i \nu W_i' \cup \mathcal{T}.$$

From the Kirby diagrams constructed in Chapter 13, it follows that C is contained in D^4, otherwise known as a 0-handle, such that the framed boundary of $\mathring{A}$ is an unknot in $\partial D^4 = S^3$. Think of $\overline{D}$ as a 2-handle attached to this unknot with framing $+1$. Then $C \cup \overline{D}$ embeds in the twisted disc bundle $S^2 \widetilde{\times} D^2$ with Euler number $+1$.

Consequently, $\overline{D} \cup C$ can be embedded within the 2-skeleton of the standard handle decomposition of $\mathbb{CP}^2$, which recall consists of a 0-handle, a 2-handle attached along a $+1$-framed unknot, and a 4-handle. In particular, $\overline{D}$ is identified with the 0-handle and C is contained within the 2-handle. By an abuse of notation, we say that the space $\overline{D} \cup C$ is contained in both M and $\mathbb{CP}^2$.

By Theorem 18.9, the tower $\mathcal{T}$ contains a mutually disjoint collection of skyscrapers with the same attaching region and by Theorem 27.1, every skyscraper is homeomorphic to a 2-handle with the same attaching region. As a result, C contains a flat, embedded disc such that its union with a disc bounded by the attaching circle in $\overline{D}$ yields a locally flat, embedded sphere $\Sigma \subseteq \overline{D} \cup C \subseteq \mathbb{CP}^2$ generating $H_2(\mathbb{CP}^2)$. Let $\mathcal{R}$ denote the complement $\mathbb{CP}^2 \smallsetminus \Sigma$.

We claim that $\mathcal{R}$ is homeomorphic to $\mathbb{R}^4$. This follows from Section 21.3.1 since $\mathcal{R}$ is proper homotopy equivalent to $\mathbb{R}^4$. In particular, to see that $\mathcal{R}$ is simply connected at infinity, we need that skyscrapers are homeomorphic to 2-handles and not just that they contain properly embedded discs. This implies that the end of $\mathcal{R}$ is topologically $S^3 \times \mathbb{R}$.

Next we show that $\mathcal{R}$ is not diffeomorphic to $\mathbb{R}^4$. Suppose for a contradiction that it is. Then there exist arbitrarily large, round, smooth, separating 3-spheres in $\mathcal{R}$. In particular, there is a smooth 3-sphere S in $\mathcal{R}$ such that the compact set $\mathbb{CP}^2 \smallsetminus \operatorname{Int}(\overline{D} \cup C)$ is contained within the ball bounded by S. In other words $S \subseteq \overline{D} \cup C$. Since $\overline{D} \cup C$ was originally embedded in M, we have $S \subseteq M \smallsetminus \Sigma$. Cut M along S, discard the component containing Σ, and glue in a 4-ball. Since we have cut and pasted along a smooth sphere, the result is a smooth manifold. Moreover, it is a closed, smooth 4-manifold with intersection form $-E_8 \oplus [-1]$, which contradicts Donaldson's Theorem A. This completes the proof.

See [**Gom83**; **Kir89**, Theorem XIV.1; **GS99**, Theorem 9.4.3] for similar proofs using Casson handles. □

The exotic $\mathbb{R}^4$ constructed above contains a smooth, compact submanifold that cannot be embedded in $\mathbb{R}^4$ with its standard smooth structure (see [**GS99**, Theorem 9.4.3]). Such an exotic $\mathbb{R}^4$ is called a *large exotic* $\mathbb{R}^4$. Exotic $\mathbb{R}^4$s that do not have such submanifolds are said to be *small*, and are constructed next.

PROOF USING THE SMOOTH FAILURE OF THE h-COBORDISM THEOREM. This proof follows [**Cas86**, Lecture III; **Kir89**, Theorem XIV.3; **DMF92**, Theorem 3.1; **GS99**, Theorem 9.3.1], once again modified to use skyscrapers.

In order to leverage the smooth failure of the h-cobordism theorem, we must begin with a smooth h-cobordism which is topologically, but not smoothly, a product. Let N be such an h-cobordism between closed, simply connected 4-manifolds M_0 and M_1, which are homeomorphic but not diffeomorphic.

We show first that this apparatus exists. Recall from Section 21.2.2 that there exist pairs of closed, smooth 4-manifolds M_0 and M_1, e.g. $\mathbb{CP}^2 \# 9\overline{\mathbb{CP}^2}$ and the Dolgachev surface, which are homeomorphic but not diffeomorphic. By [**Wal64**], two smooth, closed, simply connected 4-manifolds with isometric intersection forms are h-cobordant. Any such h-cobordism will suffice for the upcoming proof: it must be a topological product by the topological h-cobordism theorem (Theorem 20.2) but it cannot be a smooth product since M_0 and M_1 are not diffeomorphic.

As in Chapters 1 and 20, find a handle decomposition of N built on M_0 consisting of only 2-handles and 3-handles. Let $M_{1/2}$ denote the 4-manifold obtained after adding the 2-handles to $M_0 \times [0,1]$. As before, $M_{1/2}$ is simply connected and contains the attaching spheres $\{T_i\}$ of the 3-handles of N and the belt spheres $\{S_i\}$ of the 2-handles, such that $\lambda(S_i, T_j) = \delta_{ij}$. Assume henceforth that there is a single belt sphere S and a single attaching sphere T, intersecting algebraically once but geometrically $2n+1$ times.

REMARK 21.6. The restriction to a single 2-/3-handle pair is for simplicity of exposition and the proof does not depend on it. However, we observe here that there do exist such smooth h-cobordisms with non-diffeomorphic boundary components, that are built using a single 2-/3-handle pair. The first example was found by Akbulut in [**Akb91**] between blowups of the $K3$ surface and its logarithmic 0-transform. Further nontrivial h-cobordisms between blowups of elliptic surfaces and their logarithmic 0-transforms were constructed by Bižaca and Gompf in [**BG96**].

As in Chapter 20, at the expense of finger moves between S and T it can be arranged that $S \cup T$ is π_1-negligible in $M_{1/2}$. This was done by constructing framed, immersed spheres $\widehat{S}$ and $\widehat{T}$ so that S and $\widehat{S}$, and respectively T and $\widehat{T}$ are geometrically transverse and moreover, $S \cap \widehat{T} = \widehat{S} \cap T = \emptyset$.

Let $p_0, \dots, p_{2n}$ denote the points of intersection of S and T, and assume, modulo changing the orientation of T, that p_i has sign $(-1)^i$ for each i. Let M' denote the complement in $M_{1/2}$ of a small, connected regular neighbourhood $\nu(S \cup T)$ of $S \cup T$. Locate mutually disjoint, embedded Whitney circles γ_i for the intersection points $\{p_{i-1}, p_i\}$ in $\partial M'$, for $i = 1, \dots, 2n$. Since $S \cup T$ is π_1-negligible in $M_{1/2}$, $\pi_1(M') \cong \pi_1(M_{1/2})$ is trivial. Thus there exist framed, immersed, transversely intersecting discs $\{C_i\}_{i=1}^{2n}$ in M' bounded by the $\{\gamma_i\}$.

Next, we obtain transverse spheres for $\{C_i\}$. These will be constructed from Clifford tori surgered using meridional discs as usual. Let Σ_i be the surgered Clifford torus at p_{i-1} for $i = 1, \dots, 2n$ so that we have $\lambda(\Sigma_1, C_i) = 1$ if and only if $i = 1$, as well as $\lambda(\Sigma_i, C_j) = 1$ if and only if $j = i$ or $i-1$, for all $i = 2, \dots, 2n$. The result $R_2 := \Sigma_2 - \Sigma_1$ of tubing together Σ_2 and Σ_1, the latter with reversed orientation, satisfies $\lambda(R_2, C_2) = 1$ and $\lambda(R_2, C_1) = 0$. In general, define $R_i := \sum_{k=1}^{i} (-1)^{k-1} \Sigma_{i+1-k} = \Sigma_i - \Sigma_{i-1} + \dots \pm \Sigma_1$ for each $i = 1, \dots, 2n$. Then the family $\{R_i\}$ is algebraically transverse to $\{C_i\}$. Then apply the techniques of Part II to construct mutually disjoint skyscrapers $\{\widehat{\mathcal{S}}_i\}_{i=1}^{2n}$ bounded by the $\{\gamma_i\}$.

Let X denote the union $\nu(S \cup T) \cup \{\widehat{\mathcal{S}}_i\}$. Observe that since skyscrapers are standard, the space X is homeomorphic to $S^2 \times S^2 \smallsetminus \{*\}$. Further, surgery on X along S, realised by attaching the 2-handle of N, turned upside down, changes X to a space R_0, which is homeomorphic to $\mathbb{R}^4$ since surgery on $S^2 \times \{1\}$ in $S^2 \times S^2 \smallsetminus \{*\}$ gives $\mathbb{R}^4$, and S is topologically isotopic to $S^2 \times \{1\}$. Similarly, surgery on X along T, realised by attaching the 3-handle of N, turns X to a space R_1 homeomorphic to $\mathbb{R}^4$. Let U denote the union of the 5-dimensional handles arising as the trace of these surgeries. The resulting cobordism U is by construction an h-cobordism between R_0 and R_1.

Choose a Morse function $F \colon U \to [0,1]$ with $F(R_i) = i$ for $i = 0, 1$ and with critical points of index 2 and 3 at the barycentres of the 5-dimensional handles. Let ξ be a gradient-like vector field subordinate to F. Let $K \subseteq U$ denote the compact set formed as the union of the ascending and descending manifolds of the index 2 and the index 3 critical points. The complements $N \smallsetminus K$ and $U \smallsetminus K$ are both smooth products because neither contains any critical points.

Next we show that U embeds in $S^4 \times [0,1]$ with $R_i \subseteq S^4 \times \{i\}$ for each $i = 0, 1$ and such that the product structure on $U \smallsetminus K$ extends to $S^4 \times [0,1] \smallsetminus K$. To achieve this, start with $S^4 \times [0,1]$ and add a cancelling 2-handle/3-handle pair to $S^4 \times \{1\}$. Then perform n finger moves in the middle level, altering the attaching map of the 3-handle, to add cancelling pairs of intersections between the belt sphere and attaching sphere of the 2-handle and 3-handle respectively. Next add trivial tubes to the embedded Whitney and dual-Whitney discs, building the first surfaces stages of the $\{\widehat{\mathcal{S}}_i\}$. Repeat this process for each level of the $\{\widehat{\mathcal{S}}_i\}$, adding trivial local cusps and performing fingers moves for the disc stages, and tubing for the surface stages. Perform this process *with squeezing*, as in Chapter 18. The end result of this infinite process is an embedding of the open skyscrapers, and since we have performed the embedding with squeezing, the closures are embeddings of the compactified $\{\widehat{\mathcal{S}}_i\}$.

By hypothesis, the manifolds M_0 and M_1 are not diffeomorphic. We now show that this implies that R_i is not diffeomorphic to $\mathbb{R}^4$ with its standard smooth structure for each i. We will show the contrapositive.

Suppose R_0 is diffeomorphic to $\mathbb{R}^4$. Then there is a smooth, closed 4-ball $D_0 \subseteq R_0$ with $K \cap R_0 \subseteq \operatorname{Int} D_0$. The product structure on $N \smallsetminus K$ maps ∂D_0 to a smooth 3-sphere in R_1 bounding a compact submanifold D_1 with $K \cap R_1 \subseteq D_1 \subseteq R_1$, so that $M_1 \smallsetminus \operatorname{Int} D_1$ is diffeomorphic to $M_0 \smallsetminus \operatorname{Int} D_0$ by the product structure on $N \smallsetminus K$. Use the embedding of U in $S^4 \times [0,1]$ to see that $S^4 \times \{0\} \smallsetminus \operatorname{Int} D_0$ is diffeomorphic to $S^4 \times \{1\} \smallsetminus \operatorname{Int} D_1$.

Since D_0 is diffeomorphic to D^4, so is $S^4 \times \{0\} \smallsetminus \operatorname{Int} D_0$, since the closure of the complement of a smooth ball in S^4 is itself diffeomorphic to D^4. This follows from the fact that any two smooth embeddings of D^4 in a smooth 4-manifold are ambiently isotopic, up to orientation, by Palais's theorem [**RS72**, Theorem 3.34]. Then we have

$$D^4 \cong S^4 \times \{0\} \smallsetminus \operatorname{Int} D_0 \cong S^4 \times \{1\} \smallsetminus \operatorname{Int} D_1$$

As the closure of the complement of a smooth D^4 in S^4, the space D_1 is also diffeomorphic to D^4.

Consequently, the manifolds M_0 and M_1 are obtained from the diffeomorphic spaces $M_0 \smallsetminus \operatorname{Int} D_0$ and $M_1 \smallsetminus \operatorname{Int} D_1$ by adding the 4-handles D_0 and D_1 respectively. Thus, M_0 and M_1 are diffeomorphic.

It then follows from the fact that M_0 and M_1 are not diffeomorphic, that R_0 is not diffeomorphic to $\mathbb{R}^4$, so R_0 is an exotic $\mathbb{R}^4$ as asserted. □

We point out that as part of the proof above we showed that the exotic copies of $\mathbb{R}^4$, namely R_0 and R_1, are smooth submanifolds of $\mathbb{R}^4$ with its standard smooth structure. Thus they are *small exotic* $\mathbb{R}^4$*s*, as promised. According to [**Kir89**, Addendum, p. 101], the spaces R_0 and R_1 are diffeomorphic to each other.

A further construction of exotic copies of $\mathbb{R}^4$ was given by De Michelis and Freedman in [**DMF92**], consisting of adding a Casson handle to a ribbon disc complement along the meridian, and then removing the boundary. That the resulting spaces are homeomorphic to $\mathbb{R}^4$ follows from the fact that Casson handles are homeomorphic to $D^2 \times \mathring{D}^2$ relative to the attaching region. In [**DMF92**], techniques from gauge theory are used to demonstrate that the resulting spaces may be exotic copies of $\mathbb{R}^4$. In [**BG96**], Bižaca and Gompf improved these results, in particular showing that the simplest Casson handle, with a single kink at each stage, all with positive sign, may be used to construct an exotic $\mathbb{R}^4$ as above. As a byproduct, they produced a simple, albeit necessarily infinite, Kirby diagram for an exotic $\mathbb{R}^4$.

There is also a construction of a large exotic $\mathbb{R}^4$ via the existence of topologically slice knots that are not smoothly slice. The argument, due to [**Gom85**], was given in Chapter 1 (Theorem 1.15), and uses the result of Quinn that connected, noncompact 4-manifolds are smoothable (see Section 21.4).

Gompf showed in [**Gom85**] that the collection of exotic smooth structures on $\mathbb{R}^4$ is at least countably infinite. Taubes [**Tau87**] showed that there are uncountably many large exotic $\mathbb{R}^4$s, using his noncompact version of Donaldson's theorem. Later De Michelis and Freedman used similar techniques to show that there are uncountably many small exotic $\mathbb{R}^4$s [**DMF92**]. It was also shown by Freedman and Taylor [**FT86**] that there exists a smooth structure on $\mathbb{R}^4$ such that every other smoothing of $\mathbb{R}^4$ can be smoothly embedded therein. See [**GS99**, Chapter 9; **Kir89**, Chapter XIV] for further details on exotic $\mathbb{R}^4$s.

21.6.6. Computation of the 4-dimensional bordism group. The calculation of the 4-dimensional topological oriented bordism group is due to Hsu [**Hsu87**].

Theorem ($\Omega_4^{\mathrm{STOP}} \cong \mathbb{Z} \oplus \mathbb{Z}/2$). *The oriented 4-dimensional topological bordism group* $\Omega_4^{\mathrm{STOP}} \cong \mathbb{Z}\oplus\mathbb{Z}/2$*, with* $[M] \mapsto (\sigma(M), \mathrm{ks}(M))$*, where* $\sigma(M) \in \mathbb{Z}$ *is the signature of the intersection form and* $\mathrm{ks}(M) \in \mathbb{Z}/2$ *is the Kirby-Siebenmann invariant.*

Hsu's proof uses the classification of closed, simply connected 4-manifolds (Section 21.6.1), and the three fundamental tools of the immersion lemma, transversality, and normal bundles (Sections 21.4.10, 21.4.9, and 21.4.8). We give a different proof here that appeals to us. We use the sum stable smoothing theorem (Section 21.4.5), the existence of the E_8-manifold (Section 21.3.2), and some facts about smooth bordism groups.

Recall that the Kirby-Siebenmann invariant $\mathrm{ks}(M) \in H^4(M;\mathbb{Z}/2)$ of a closed, connected topological 4-manifold M is the obstruction for the stable tangent microbundle classifying map $\tau_M \colon M \to BTOP$ to admit a lift to $\tau'_M \colon M \to BPL$. For $M = \sqcup^i M_i$ a union of connected components we define $\mathrm{ks}(M) := \sum_i \mathrm{ks}(M_i)$. This defines a homomorphism $\Omega_4^{\mathrm{STOP}} \to \mathbb{Z}/2$, which is surjective due to the existence of the E_8-manifold, and which vanishes for 4-manifolds bordant to a smooth manifold. See [**FNOP19**, Section 8] for details on the Kirby-Siebenmann invariant of 4-manifolds.

Proof. By the sum stable smoothing theorem [**FQ90**, Theorem 8.6] (Section 21.4.5), a closed 4-manifold with vanishing Kirby-Siebenmann invariant admits a smooth structure after connected sum with copies of $S^2 \times S^2$. Since $S^2 \times S^2$ is null-bordant, and since the smooth 4-dimensional oriented bordism group is $\Omega_4^{SO} \cong \mathbb{Z}$, detected by the signature, we have a short exact sequence

$$0 \to \Omega_4^{SO} \to \Omega_4^{\mathrm{STOP}} \xrightarrow{\mathrm{ks}} \mathbb{Z}/2 \to 0.$$

The signature provides a splitting homomorphism, so

$$\Omega_4^{\mathrm{STOP}} \cong \mathbb{Z} \oplus \mathbb{Z}/2$$

as claimed. □

CHAPTER 22

Surgery theory and the classification of closed, simply connected 4-manifolds

Patrick Orson, Mark Powell, and Arunima Ray

For X a 4-dimensional Poincaré complex with $\pi_1(X)$ good, we will explain the terms and the maps in the (simple) surgery sequence

$$L_5^s(\mathbb{Z}[\pi_1(X)]) \to \mathcal{S}^s(X) \xrightarrow{\rho} \mathcal{N}(X) \xrightarrow{\sigma} L_4^s(\mathbb{Z}[\pi_1(X)])$$

in more detail. Since the standard surgery theory literature focusses on dimension at least five, and the treatment in [**FQ90**] assumes prior knowledge of surgery theory, we give an account that requires less background, and explain exactly where the sphere embedding theorem is used, purely focussing on the 4-dimensional case. The category preserving sphere embedding theorem will be used both to show exactness at $\mathcal{N}(X)$ as well as to define the action of $L_5^s(\mathbb{Z}[\pi_1(X)])$ on the simple structure set $\mathcal{S}^s(X)$. The version of the sphere embedding theorem proved in this book, where the ambient manifold is required to be smooth, could be used to prove corresponding weaker results, with smooth input and topological output. For results lying fully in the topological category, we need the category preserving sphere embedding theorem, described in the previous chapter, whose proof requires topological transversality, the immersion lemma, and the existence of normal bundles for locally flat submanifolds of a topological 4-manifold. See Chapter 21 for more details. In the present and following chapter, we use these results, pointing out when and why they are needed.

In the second part of this chapter, we explain the use of the surgery sequence and the s-cobordism theorem in the classification of closed, simply connected topological 4-manifolds, up to homeomorphism.

The existence and exactness of the surgery sequence refers to the following two statements for Poincaré complexes X with $\pi_1(X)$ good. These statements will be made precise in this chapter.

(i) Let (M, f, ξ, b) be a degree one normal map in $\mathcal{N}(X)$. Then (M, f, ξ, b) is equivalent (normally bordant) to a degree one normal map (M', f', ξ', b') with f' a simple homotopy equivalence if and only if $\sigma(M, f, \xi, b) = 0 \in L_4^s(\mathbb{Z}[\pi_1(X)])$. In particular, if there exists (M, f, ξ, b) with $\sigma(M, f, \xi, b) = 0$, then $\mathcal{S}^s(X) \neq \emptyset$, and the surgery sequence for X exists and is exact at $\mathcal{N}(X)$ as a sequence of pointed sets.

(ii) Wall realisation determines an action of the group $L_5^s(\mathbb{Z}[\pi_1(X)])$ on $\mathcal{S}^s(X)$, the orbits of which coincide with the point preimages

$$\big\{\rho^{-1}([(M, f, \xi, b)]) \mid [(M, f, \xi, b)] \in \mathcal{N}(X)\big\}.$$

In other words, the action of $L_5^s(\mathbb{Z}[\pi_1(X)])$ exists, and the surgery sequence is exact at the structure set $\mathcal{S}^s(X)$.

This chapter is not intended as a substitute for a text on surgery theory, and the reader who wants to learn more is recommended to consult one of these, for example [**Wal99**, **Ran02**, **Lüc02**].

22.1. The surgery sequence

Much of this section is a review of the basics of surgery theory from the perspective of 4-manifolds. The reader might also refer to [**KT01**, **Wal99**]. As indicated by the decorations in the sequence above, we are discussing a slightly different surgery sequence than in Chapter 1. The version there is up to h-cobordism, while here we work with s-cobordisms. The former is easier to define, whence its appearance in the introduction, but the latter is the form required for potential applications to non-simply connected manifolds, so we have chosen to include this level of generality.

In the surgery sequence, and the rest of this chapter, X is a 4-dimensional *Poincaré complex*, that is a finite CW complex, equipped with an orientation character $w\colon \pi_1(X) \to \mathbb{Z}/2$ and a *fundamental class* $[X] \in H_4(X;\mathbb{Z}^w)$ such that cap product with $[X]$ induces a simple chain equivalence

$$- \cap [X]\colon C^{4-*}(X;\mathbb{Z}[\pi_1(X)]^w) \xrightarrow{\simeq} C_*(X;\mathbb{Z}[\pi_1(X)]).$$

This is what is meant at the beginning of Chapter 1, when we say that X satisfies Poincaré duality. Observe that for a closed topological manifold M equipped with a nontrivial orientation character there is a canonical choice of a fundamental class. For an oriented topological manifold, the orientation character is of course trivial, and there are two choices of fundamental class per connected component. We will often assume that $\pi_1(X)$ is good and we will point out explicitly where this hypothesis is needed.

Every compact topological manifold M embeds in high dimensional Euclidean space (see, for example, [**Hat02**, Corollary A.9]). Indeed, it is shown in [**KS77**, Theorem 5.13, Essay III] that there exists an embedding with a tubular neighbourhood, which is a finite CW complex homotopy equivalent to M (see also [**FNOP19**, Theorem 4.5]), and moreover this process results in a Poincaré complex. Chapman's theorem [**Cha74**] states that any two CW structures on a compact topological space are simple homotopy equivalent. Thus a compact topological manifold determines a Poincaré complex, unique up to simple homotopy equivalence. For the rest of this chapter, we will assume every compact topological manifold M comes with a choice of Poincaré complex structure. This includes in particular a fundamental class.

Given a Poincaré complex X, the principal aim of the surgery sequence is to compute the *simple structure set* $\mathcal{S}^s(X)$, which by definition consists of equivalence classes of pairs (M,f), where M is a closed topological 4-manifold and $f\colon M \to X$ is a simple homotopy equivalence, respecting fundamental classes. The word simple is meaningful here since M is equipped with a choice of Poincaré complex structure, which is unique up to simple homotopy equivalence. The equivalence relation is defined by setting $(M,f) \sim (M',f')$ when there exists a cobordism $F\colon W \to X \times [0,1]$ with boundary $\partial(W,F) = -(M,f) \sqcup (M',f')$ such that F is a simple homotopy equivalence. In this case, we say that (M,f) and (M',f') are *s-cobordant* over X. Since such a cobordism W is in particular an s-cobordism, then if $\pi_1(X)$ is a good group the category preserving 5-dimensional s-cobordism theorem (Section 21.5) implies that this equivalence relation is the same as that of homeomorphism over X.

In the typical surgery programme for classifying closed (oriented) manifolds up to homeomorphism within a fixed simple homotopy equivalence class, one computes the quotient of $\mathcal{S}^s(X)$ by simple self-homotopy equivalences of X which respect the fundamental class. This is sufficient since for a fixed closed, topological 4-manifold M, if we have simple homotopy equivalences $f, f' \colon M \to X$, the map $f^{-1} \circ f' \colon M \to M$ is a simple self-homotopy equivalence and given any simple self-homotopy equivalence $\phi \colon M \to M$ and simple homotopy equivalence $f \colon M \to X$, the map $f \circ \phi \colon M \to X$ is a simple homotopy equivalence. To obtain the unoriented classification, one then factors out by the choice of fundamental class. See [**CM19**] for more on the passage from the structure set to the set of manifolds up to homeomorphism.

Note that for a given X, there might not be any topological 4-manifold simple homotopy equivalent to X, in which case $\mathcal{S}^s(X)$ is empty and there is no surgery sequence. If $\mathcal{S}^s(X)$ is nonempty, then one must fix a choice of topological manifold M with a simple homotopy equivalence $f \colon M \to X$ as the distinguished point in $\mathcal{S}^s(X)$. If X is itself a topological manifold, (X, Id) is the distinguished point.

22.1.1. Normal maps. Define $G(k)$ to be the monoid of self-homotopy equivalences of S^{k-1}. The space $BG(k)$ is the classifying space for fibrations with fibres homotopy equivalent to S^{k-1} and fibre automorphisms given by self-homotopy equivalences. Define BG to be the colimit of $\{BG(k)\}$ over all k. As before, define the space $BTOP(k)$ to be the classifying space for $\mathbb{R}^k$ fibre bundles with structure set $\mathrm{Homeo}_0(\mathbb{R}^k)$, the set of homeomorphisms of $\mathbb{R}^k$ that fix the origin, and define $BTOP$ to be the colimit of $\{BTOP(k)\}$ over all k. There is a forgetful map $BTOP \to BG$ defined by restricting to $\mathbb{R}^k \smallsetminus \{0\}$ on each $BTOP(k)$. Henceforth we refer to $\mathbb{R}^k$ fibre bundles with structure set $\mathrm{Homeo}_0(\mathbb{R}^k)$ as (topological) vector bundles.

A Poincaré complex X comes equipped with a canonical stable spherical fibration classified by (the homotopy class of) a map $X \to BG$, called the *Spivak normal fibration* [**Spi67**]. A topological manifold M comes equipped with a canonical stable (topological) vector bundle classified by a map $M \to BTOP$, called the *stable normal bundle* of M, denoted by ν_M.

The set of *normal maps* $\mathcal{N}(X)$ is the set of equivalence classes of quadruples (M, f, ξ, b) where $f \colon M \to X$ is a degree one map from a closed, topological 4-manifold M to X mapping the fundamental class $[M]$ to $[X]$, together with a stable (topological) vector bundle $\xi \to X$ and a bundle map $b \colon \nu_M \to \xi$. In other words, we have the following diagram

$$\begin{array}{ccc} \nu_M & \xrightarrow{b} & \xi \\ \downarrow & & \downarrow \\ M & \xrightarrow{f} & X. \end{array}$$

Since normal maps are often the input to the surgery programme, we sometimes refer to a normal map as a *surgery problem*.

Two such quadruples (M, f, ξ, b) and (M', f', ξ', b') are said to be equivalent if they are cobordant over X, that is if there exists a quadruple (W, F, Ξ, B) consisting of a cobordism $F \colon W \to X \times [0,1]$ with boundary $\partial(W, F) = -(M, f) \sqcup (M', f')$ such that the fundamental class $[W]$ maps to $[X \times [0,1]] \in H_5(X \times [0,1], X \times \{0,1\}; \mathbb{Z}^w)$, a stable vector bundle $\Xi \to X \times [0,1]$, and a stable bundle map $B \colon \nu_W \to \Xi$ such that the bundle data extend the given bundle data (ξ, b) and (ξ', b') on M and M' respectively.

Remark 22.1. If the Spivak normal fibration of a Poincaré complex X lifts to $X \to BTOP$ we say it has a topological vector bundle reduction. By a result of Land [**Lan17**, Theorem 3.3], based on [**GS65**] (see also [**Ham18**]), the Spivak normal fibration of an oriented 4-dimensional Poincaré complex always admits a topological vector bundle reduction.

If a Poincaré complex X admits a topological vector bundle reduction, then there exists a degree one normal map (M, f, ξ, b) to X, with respect to the given reduction [**Ran02**, Theorem 9.42]. This uses topological transversality, which is discussed further in Sections 21.4.9 and 21.5.1.

Now we define the map $\mathcal{S}^s(X) \to \mathcal{N}(X)$. Let (M, f) represent an element of $\mathcal{S}^s(X)$. Choose a homotopy inverse $g\colon X \to M$ for $f\colon M \to X$, and define $\xi := g^*(\nu_M)$ to be the pullback of the stable normal bundle of M. A stable bundle map $b\colon \nu_M \to \xi$ is equivalent to an isomorphism $f^*(\xi) = f^* \circ g^*(\nu_M) = (g \circ f)^* \nu_M \cong \nu_M$, which we obtain from a homotopy $h\colon g \circ f \sim \mathrm{Id}\colon M \to M$. The image of (M, f) in $\mathcal{N}(X)$ is defined to be (M, f, ξ, b) and the distinguished point of $\mathcal{N}(X)$ is by definition the image of the chosen distinguished point of $\mathcal{S}^s(X)$. A different choice of homotopy $h'\colon g \circ f \sim \mathrm{Id}$ determines a different stable bundle map b'. However, as $h, h'\colon M \times [0,1] \to M$ are both the identity on one end they are each homotopic to the projection $M \times [0,1] \to M$. Stacking these, there is a homotopy $\bar{h}\colon (M \times [0,1]) \times [0,1] \to M$ from h to h'. This homotopy can be used to construct a degree one normal bordism $(M \times [0,1], f \times \mathrm{Id}, (g \times \mathrm{Id})^* \nu_{M \times [0,1]}, B)$ from (M, f, ξ, b) to (M, f, ξ, b'), proving that the normal bordism class of (M, f, ξ, b) does not depend on the choice of h (and thus of b). Moreover, if g' is another choice of homotopy inverse for f, then there is a homotopy $g' = \mathrm{Id} \circ g' \sim g \circ f \circ g' \sim g \circ \mathrm{Id} = g$, which can similarly be used to show that the normal bordism class does not depend on the choice of g, and thus of ξ.

To complete the argument that the map is well defined one must also show that (M, f) and (M', f'), representing equal elements in $\mathcal{S}^s(X)$, are mapped to the same element in $\mathcal{N}(X)$. We show this, in a slightly more general context, in Lemma 23.27.

22.1.2. *L*-groups. Let $^-$ denote the involution on $\mathbb{Z}[\pi_1(X)]$ generated by sending $g \mapsto w(g)g^{-1}$ for every $g \in \pi_1(X)$, where as before w denotes the orientation character. Recall that a form $\lambda\colon P \times P \to \mathbb{Z}[\pi_1(X)]$ on a finitely generated, free, based $\mathbb{Z}[\pi_1(X)]$-module P is said to be

(1) *sesquilinear* if $\lambda(ra, sb) = r \cdot \lambda(a, b) \cdot \overline{s}$, for all $r, s \in \mathbb{Z}[\pi_1(X)]$ and $a, b \in P$;
(2) *hermitian* if $\lambda(a, b) = \overline{\lambda(b, a)}$, for all $a, b \in P$;
(3) *nonsingular* if the adjoint map $\lambda^{ad}\colon P \to P^*$ sending $a \mapsto \lambda(a, -)$ is an isomorphism; and
(4) *simple* if λ^{ad} has vanishing Whitehead torsion with respect to the preferred basis of P.

A *quadratic enhancement* of λ is a function $\mu\colon P \to \mathbb{Z}[\pi_1(X)]/g \sim \overline{g}$ so that

(1) $\mu(r \cdot a) = r\mu(a)\overline{r}$, for all $r \in \mathbb{Z}[\pi_1(X)]$ and $a \in P$;
(2) $\mu(a) + \overline{\mu(a)} = \lambda(a, a)$, for all $a \in P$; and
(3) $\mu(a + b) - \mu(a) - \mu(b) = \mathrm{pr}(\lambda(a, b))$ for all $a, b \in P$, where pr is the projection map.

A triple (P, λ, μ) satisgfying all the properties above is called a *nonsingular quadratic form*. For Q a finitely generated, free, based $\mathbb{Z}[\pi_1(X)]$-module, we have a form λ on $Q \oplus Q^*$ given by

$$\begin{pmatrix} 0 & \mathrm{ev} \\ \mathrm{ev} & 0 \end{pmatrix}$$

where ev denotes either evaluation or the canonical identification $Q \cong Q^{**}$ followed by evaluation. In other words, for $(p, f), (q, g) \in Q \oplus Q^*$, define $\lambda((p, f), (q, g)) := f(q) + g(p)$. Use the quadratic enhancement for λ given by setting $\mu(q) = 0 = \mu(q^*)$ for every $q \in Q$, $q^* \in Q^*$. The nonsingular quadratic form $H(Q) := (Q \oplus Q^*, \lambda, \mu)$ is called the *standard hyperbolic form* on Q.

The sum of two nonsingular quadratic forms is constructed by taking the direct sum of each element of the triple. Two nonsingular quadratic forms (P, λ, μ) and

(P', λ', μ') are said to be isometric if there is an isomorphism $P \xrightarrow{\cong} P'$ inducing isometries of λ and λ' as well as μ and μ'.

The *L-group* $L_4^s(\mathbb{Z}[\pi_1(X)])$ is defined to be the set of nonsingular quadratic forms, modulo the equivalence relation generated by declaring two nonsingular quadratic forms (P, λ, μ) and (P', λ', μ') to be equivalent if there are finitely generated, free, based $\mathbb{Z}[\pi_1(X)]$-modules Q and Q' and an isometry

$$(P, \lambda, \mu) \oplus H(Q) \cong (P', \lambda', \mu') \oplus H(Q')$$

such that the underlying isomorphism of based modules has vanishing Whitehead torsion. In other words, $L_4^s(\mathbb{Z}[\pi_1(X)])$ consists of equivalence classes of sesquilinear, hermitian, nonsingular simple forms $\lambda \colon P \times P \to \mathbb{Z}[\pi_1(X)]$ on a finitely generated, free, based $\mathbb{Z}[\pi_1(X)]$-module P, together with a quadratic enhancement $\mu \colon P \to \mathbb{Z}[\pi_1(X)]/g \sim \overline{g}$. In this group, the inverse of (P, λ, μ) is $(P, -\lambda, -\mu)$ and the zero element is the class of the hyperbolic forms, which is also the distinguished element of $L_4^s(\mathbb{Z}[\pi_1(X)])$ thought of as a pointed set.

A free, half-rank submodule $i_L \colon L \to P$ of a nonsingular quadratic form (P, λ, μ) is known as a *lagrangian* if both λ and μ vanish on L. A lagrangian determines a short exact sequence

$$0 \longrightarrow L \xrightarrow{i_L} P \xrightarrow{(i_L)^* \circ \lambda^{ad}} L^* \longrightarrow 0$$

and a based lagrangian is called *simple* when this sequence has vanishing Whitehead torsion. It is known that a nonsingular quadratic form is isomorphic to a hyperbolic form $H(L)$, such that the underlying isomorphism of based modules has vanishing Whitehead torsion, if and only if the form admits a simple lagrangian [**Wal99**, Lemma 5.3]. Thus a nonsingular quadratic form vanishes in $L_4^s(\mathbb{Z}[\pi_1(X)])$ if and only if it admits a simple lagrangian after stabilisation by a hyperbolic form $H(Q)$ for some finitely generated, free, based $\mathbb{Z}[\pi_1(X)]$-module Q.

Next, we start to give the background needed to define the *surgery obstruction group* $L_5^s(\mathbb{Z}[\pi_1(X)])$. A *nonsingular quadratic formation* consists of a nonsingular quadratic form (P, λ, μ), as above, together with two simple lagrangians F and G. Addition of nonsingular quadratic formations is by direct sum on each of the entries in the 5-tuple. Two nonsingular quadratic formations (P, λ, μ, F, G) and $(P', \lambda', \mu', F', G')$ are isomorphic if there exists an isomorphism of modules $\theta \colon P \xrightarrow{\cong} P'$ inducing both an isometry of nonsingular quadratic forms and isomorphisms $F \cong F'$ and $G \cong G'$, such that each of these three module isomorphisms has vanishing Whitehead torsion.

Since every nonsingular quadratic form with a simple lagrangian is known to be isomorphic to a hyperbolic form, every nonsingular quadratic formation is isomorphic to $(H(P), P, G)$ for some finitely generated, free, based $\mathbb{Z}[\pi_1(X)]$-module P and for some simple lagrangian G of $H(P)$. Here we are using the fact that P is always a simple lagrangian for $H(P)$.

Equivalence of nonsingular quadratic formations is more difficult to define. We will need the following definitions.

(1) We say that two nonsingular quadratic formations, given by (P, λ, μ, F, G) and $(P', \lambda', \mu', F', G')$, are *stably isomorphic* if there are finitely generated, free, based $\mathbb{Z}[\pi_1(X)]$-modules Q, Q' such that

$$(P, \lambda, \mu, F, G) \oplus (H(Q), Q, Q^*) \cong (P', \lambda', \mu', F', G') \oplus (H(Q'), Q', (Q')^*).$$

(2) A *skew-hermitian* quadratic form (P, λ, μ) is a finitely generated, free, based $\mathbb{Z}[\pi_1(X)]$-module P with a sesquilinear form $\lambda \colon P \times P \to \mathbb{Z}[\pi_1(X)]$ that is skew-hermitian i.e. $\lambda(a, b) = -\overline{\lambda(b, a)}$, together with a quadratic enhancement $\mu \colon P \to \mathbb{Z}[\pi_1(X)]/g \sim -\overline{g}$. That is, we require that $\mu(r \cdot a) =$

$r\mu(a)\overline{r}$, $\mu(a) - \overline{\mu(a)} = \lambda(a,a)$ and $\mu(a+b) - \mu(a) - \mu(b) = \mathrm{pr}(\lambda(a,b))$ for all $p, q \in P$ and $r \in \mathbb{Z}[\pi_1(X)]$, where pr is the projection map. Note that we do not require the form to be nonsingular.

(3) Given a skew-hermitian quadratic form (P, λ, μ), we define the *boundary formation*

$$\partial(P, \lambda, \mu) := (H(P), P, \Gamma_{(P,\lambda)})$$

where $\Gamma_{(P,\lambda)} := \{(p, \lambda^{ad}(p)) \mid p \in P\} \subseteq P \oplus P^*$ is called the *graph lagrangian* of (P, λ); it does not depend on μ.

We say that two nonsingular quadratic formations (P, λ, μ, F, G) and $(P', \lambda', \mu', F', G')$ are equivalent if there exist skew-hermitian quadratic forms (Q, λ, μ) and (Q', λ', μ') such that $(P, \lambda, \mu, F, G) \oplus \partial(Q, \lambda, \mu)$ is stably isomorphic to $(P', \lambda', \mu', F', G') \oplus \partial(Q', \lambda', \mu')$.

The L-group $L_5^s(\mathbb{Z}[\pi_1(X)])$, also called the *surgery obstruction group*, consists of stable isomorphism classes of nonsingular quadratic formations, modulo boundary formations. The trivial element in the group is represented by formations $(H(F), F, F^*)$ where F is a finitely generated, free, based $\mathbb{Z}[\pi_1(X)]$-module.

As indicated by the sub- and superscripts, there are versions of the L-groups with other decorations and indices. For other non-negative indices and fixed superscript, the L-groups are 4-periodic. For $n \equiv 2, 3 \mod 4$, the definitions are similar to those we have presented but with different signs. See, for example, [**Ran02**, **Wal99**]. For other decorations, the symmetric theory, and negative indices, the interested reader should consult [**HT00**] for an initial guide, with more details in, for example, [**Ran73**; **Ran80**, Section 9; **Ran81**, Section 1.10; **Ran92**].

We will also need the following definitions.

(1) Two bases for a given finitely generated, free $\mathbb{Z}[\pi_1(X)]$-module are called *simply equivalent* if the change of basis matrix has vanishing Whitehead torsion.
(2) Two bases $\mathcal{B}_1$ and $\mathcal{B}_2$ for a given finitely generated, free $\mathbb{Z}[\pi_1(X)]$-module P are said to be *stably simply equivalent* if $\mathcal{B}_1$ and $\mathcal{B}_2$ can be extended to bases $\mathcal{B}_1'$ and $\mathcal{B}_2'$ for a stabilisation of P by a free $\mathbb{Z}[\pi_1(X)]$-module, such that $\mathcal{B}_1'$ and $\mathcal{B}_2'$ are simply equivalent.
(3) A basis for a stabilisation of a finitely generated, free $\mathbb{Z}[\pi_1(X)]$-module P is called a *stable basis* for P.

22.1.3. The surgery obstruction map. Consider a quadruple (M, f, ξ, b) representing an element of $\mathcal{N}(X)$. Since f is degree one, M is connected and $\pi_1(f)\colon \pi_1(M) \to \pi_1(X)$ is surjective. Perform surgery on classes in $\ker(\pi_1(f))$ to alter (M, f) so that $\pi_1(f)$ is an isomorphism. More precisely, surgery produces a normal bordism, by adding handles along embedded curves representing generators of $\ker(\pi_1(f))$, from (M, f, ξ, b) to some (M', f', ξ', f') where $\pi_1(f')$ is an isomorphism. But, as is customary, we will abuse notation and keep using the same symbols. In higher ambient dimensions, we would continue this process up until the middle dimension, so it is called *surgery below the middle dimension*.

We now have (M, f, ξ, b) such that f induces an isomorphism on fundamental groups. By Whitehead's theorem and Poincaré duality, since X is 4-dimensional, the sole obstruction to f being a homotopy equivalence is the module

$$\ker(\pi_2(f)) \cong K_2(f) := \ker(H_2(f)\colon H_2(M; \mathbb{Z}[\pi_1(X)]) \to H_2(X; \mathbb{Z}[\pi_1(X)])).$$

The submodule $K_2(f)$ is called the *surgery kernel* of f. Above we used the Hurewicz theorem to pass from homotopy groups to the homology groups of the universal covers. The intersection form of M restricts to an even, nonsingular, sesquilinear, hermitian form λ on the finitely generated, stably free $\mathbb{Z}[\pi_1(X)]$-module $K_2(f)$, which is known to have a preferred simple equivalence class of stable basis (see the following paragraph). Perform surgeries on trivial circles in M to add a hyperbolic form to $K_2(f)$ and whence realise any stabilisation. Again using the same notation after the surgeries, we have that $P := K_2(f)$ is now a finitely generated, free $\mathbb{Z}[\pi_1(X)]$-module with a preferred simple equivalence class of basis. If the orientation character w is trivial on order two elements of $\pi_1(M)$, then since the form λ is even, there is a unique quadratic enhancement μ, which equals the self-intersection number of elements of $K_2(f)$. In general, the normal data determine a unique regular homotopy class of immersions for each element of $K_2(f)$, which gives rise to a quadratic enhancement $\mu\colon K_2(f) \to \mathbb{Z}[\pi_1(X)]/g \sim \overline{g}$. Thus we have obtained an element $\sigma(M, f, \xi, b) = [(K_2(f), \lambda, \mu)] \in L_4^s(\mathbb{Z}[\pi_1(X)])$, which is called the *surgery obstruction* for (M, f, ξ, b). That is, we have defined the map

$$\sigma\colon \mathcal{N}(X) \to L_4^s(\mathbb{Z}[\pi_1(X)]).$$

So far we have discussed the 4-dimensional version of the procedure given in [**Wal99**, Chapters 1, 2, and 5]. Specifically, Chapter 1 performs surgery below the middle dimension as above, while Chapter 2 shows that the surgery kernel $K_2(f)$ is finitely generated and stably free with a preferred simple equivalence class of stable basis, and that the intersection form restricts to a form on $K_2(f)$ with a quadratic enhancement. Chapter 5 of [**Wal99**] constructs the surgery obstruction and shows that it only depends on the normal bordism class of (M, f, ξ, b), so σ gives a well defined map from $\mathcal{N}(X)$ to $L_4^s(\mathbb{Z}[\pi_1(X)])$. Briefly, suppose that (M, f, ξ, b) and (M', f', ξ', b') are equivalent in $\mathcal{N}(X)$ via (W, F, Ξ, B). Assume we have performed the surgeries on trivial circles in both M and M' to stabilise $K_2(f)$ and $K_2(f')$ to a free modules. Now perform surgeries on the interior of W to make $\pi_1(F)$ an isomorphism. There is now a handle decomposition of W relative to M consisting only of k 2-handles and ℓ 3-handles, for some integers k and ℓ. In the topological category, this uses that 5-manifolds admit topological handle decompositions (see Section 21.4.11). This is sufficient to determine an isomorphism

$$(K_2(f), \lambda, \mu) \oplus H(\mathbb{Z}[\pi_1(X)]^k) \cong (K_2(f'), \lambda', \mu') \oplus H(\mathbb{Z}[\pi_1(X)]^\ell).$$

To see that there is a simple isomorphism between these forms requires more care, and for this we refer the reader to [**Wal99**, Theorem 5.6].

22.1.4. Exactness at the normal maps. We now show that the surgery sequence is exact at $\mathcal{N}(X)$. Consider the image $(M, f, \xi, b) \in \mathcal{N}(X)$ of an element $(M, f) \in \mathcal{S}^s(X)$, under the map in the surgery sequence. Since f is a homotopy equivalence, no surgery below the middle dimension is necessary, and the surgery kernel $K_2(f)$ is already trivial. Thus $\sigma(M, f, \xi, b) = 0 \in L_4^s(\mathbb{Z}[\pi_1(X)])$. This shows half of the desired exactness, namely that the image of $\mathcal{S}^s(X)$ lies in the kernel of $\sigma\colon \mathcal{N}(X) \to L_4^s(\mathbb{Z}[\pi_1(X)]$.

Now suppose that (M, f, ξ, b) lies in the kernel of σ. We will show that (M, f, ξ, b) is normally bordant to a simple homotopy equivalence. That is, after a finite sequence of surgeries below the middle dimension, including on trivial circles to realise stabilisation, we have that $K_2(f) \cong \ker(\pi_2(f))$ is a free, finitely generated $\mathbb{Z}[\pi_1(X)]$-module and the intersection form is hyperbolic. Our aim is to perform surgery on 2-dimensional homotopy classes representing a lagrangian of this hyperbolic form. This means representing a half basis of $K_2(f)$ by framed, embedded 2-spheres, and for each such embedding replacing a neighbourhood $S^2 \times D^2$ with $D^3 \times S^1$. This has the effect of killing the homotopy class represented by the core $S^2 \times \{0\}$. If the embedding has a geometrically transverse sphere, then a meridian $\{\mathrm{pt}\} \times S^1$ to the removed S^2 is null-homotopic, via the transverse sphere minus its intersection with $S^2 \times D^2$. Thus the surgery operation does not affect the fundamental group, and $\pi_1(f)$ remains an isomorphism.

Thankfully, we are in exactly the situation of the (category preserving) sphere embedding theorem. Choose any simple lagrangian P for the quadratic form $(K_2(f), \lambda, \mu)$. There is then an isomorphism from $(K_2(f), \lambda, \mu)$ to the hyperbolic form $H(P)$, such that the isomorphism on modules has vanishing Whitehead torsion. Consider classes $\{f_i\}$ generating P, and classes $\{g_i\}$ generating P^*. Then after adding local cusps, which corresponds to restricting to the preferred regular homotopy classes determined by the normal data, the sets $\{f_i\}$ and $\{g_i\}$ can be represented by framed, immersed spheres $S^2 \looparrowright M$, such that $\lambda(f_i, g_j) = \delta_{ij}$, $\lambda(f_i, f_j) = 0$, and $\mu(f_i) = 0$ for all i, j. By hypothesis, $\pi_1(M) \cong \pi_1(X)$ is good.

Then the category preserving sphere embedding theorem (Section 21.5) says that the $\{f_i\}$ are regularly homotopic to a collection of mutually disjoint, locally flat embedded spheres $\{\overline{f}_i\}$ with geometrically transverse spheres $\{\overline{g}_j\}$. The set $\{\overline{f}_i\}$ is framed since the set $\{f_i\}$ was. Use $\{\overline{f}_i\}$ as the data for surgery to construct a normal bordism from (M, f, ξ, b) to some (M', f', ξ', b') such that $f' \colon M' \to X$ induces an isomorphism on π_i for $i = 0, 1, 2$. Then, as mentioned earlier, by Poincaré duality and the Hurewicz theorem f' induces an isomorphism on all homotopy groups and is therefore a homotopy equivalence by Whitehead's theorem. Moreover, the homotopy equivalence $f' \colon M' \to X$ is simple by [**Wal99**, Theorem 5.6].

While the concept of exactness at the normal maps technically requires that the structure set be nonempty, note that the argument of this subsection can also be used to show that a structure set is nonempty, by showing that the kernel of σ is nonempty.

22.1.5. Wall realisation. Suppose that the structure set of the Poincaré complex X is nonempty. We will define an action of the surgery obstruction group $L_5^s(\mathbb{Z}[\pi_1(X)])$ on $\mathcal{S}^s(X)$. The leftmost arrow in the surgery sequence refers to this action. Exactness at the structure set means, by definition, that two elements of the structure set are in the same orbit of this action if and only if they agree when mapped to $\mathcal{N}(X)$. The action will be defined using *Wall realisation*, a process we use to geometrically realise given elements of $L_5^s(\mathbb{Z}[\pi_1(X)])$.

The definition of Wall realisation uses the sphere embedding theorem, and therefore requires that $\pi_1(X)$ be a good group. We start with a degree one normal map (M, f, ξ, b) to X, such that f is a simple homotopy equivalence, together with a given nonsingular quadratic formation. As noted earlier, every nonsingular quadratic formation is isomorphic to $(H(P), P, G)$ for some finitely generated, free, based $\mathbb{Z}[\pi_1(X)]$-module P and for some simple lagrangian G of $H(P)$. Let k be the rank of P. Perform k surgeries on trivial, embedded circles in M. The *trace* of these surgeries consists of $M \times [0,1]$ with k 5-dimensional 2-handles $D^2 \times D^3$ attached to the trivial circles. Choose framings on the circles such that this builds a cobordism W' over $X \times [0, 1/2]$ from $f \colon M \to X$ to $f' \colon M' := M \#^k S^2 \times S^2 \to X$. The surgery kernel $K_2(f) = 0$ changes to $(K_2(f'), \lambda, \mu) \cong H(P)$, the intersection form of $\#^k S^2 \times S^2$. Here the summand P is identified with the submodule generated by the spheres $S^2 \times \{\mathrm{pt}\}$.

Now we come to the second application of the category preserving sphere embedding theorem in this chapter. The lagrangian $G \subseteq P \oplus P^* \cong K_2(f' \colon M \#^k S^2 \times S^2 \to X)$ is a finitely generated, free, based submodule of $K_2(f')$. Represent the basis of G by framed, immersed spheres $\{f_1, \dots, f_k\}$ in $M \#^k S^2 \times S^2$ with $\lambda(f_i, f_j) = 0$ for all $i, j = 1, \dots, k$ and $\mu(f_i) = 0$ for $i = 1, \dots, l$. Since the intersection form on $M \#^k S^2 \times S^2$ is nonsingular, there is a collection of dual spheres $\{g_1, \dots, g_k\}$, also framed and immersed, such that $\lambda(f_i, g_j) = \delta_{i,j}$ for all i, j.

By the category preserving sphere embedding theorem, the spheres $\{f_i\}$ are regularly homotopic to a collection of mutually disjoint, locally flat embedded spheres $\{\overline{f}_i\}$ with geometrically transverse spheres $\{\overline{g}_j\}$. The spheres $\{\overline{f}_i\}$ are framed since the spheres $\{f_i\}$ were. Use the spheres $\{\overline{f}_i\}$ as the data for surgery on $M\#^k S^2 \times S^2$. The trace of this surgery is a cobordism W'' over $X \times [1/2,1]$ from $M\#^k S^2 \times S^2$ to another closed, topological 4-manifold M''. The map to X extends because we perform surgery on classes that map to null-homotopic elements of X. The second surgery kills $K_2(f')$, but in a different way than how surgery on the generators of P would kill it. The second surgery does not create new generators of the fundamental group, again due to the geometrically transverse spheres. Observe that the normal data (ξ, b) can be extended across the cobordism W, since we performed surgery on compatibly framed spheres representing relative homotopy classes. Thus we have produced a degree one normal bordism (W, F, Ξ, B) from (M, f, ξ, b) to a new degree one normal map (M'', f'', ξ'', b''). The resulting map $f''\colon M'' \to X$ is again a homotopy equivalence and moreover, a simple homotopy equivalence. This latter fact is proved in [**Wal99**, Theorem 6.5], but using an alternative definition of the groups $L_5^s(\mathbb{Z}[\pi_1(X)])$ and hence a slightly different, but equivalent, definition of the action on $\mathcal{S}^s(X)$. By construction, the degree one normal map (W, F, Ξ, B) has surgery obstruction $\sigma(W, F, \Xi, B) = [(H(P), P, G)] \in L_5^s(\mathbb{Z}[\pi_1(X)])$. We did not explain how to extract an element of $L_5^s(\mathbb{Z}[\pi_1(X)])$ from an odd dimensional surgery problem. We refer the reader instead to, for example, [**Ran02**, **Lüc02**].

22.1.6. Exactness at the structure set. Using Wall realisation we define the action of the surgery obstruction group $L_5^s(\mathbb{Z}[\pi_1(X)])$ on the structure set $\mathcal{S}^s(X)$. We start with an element of the structure set, a closed, topological 4-manifold M with a simple homotopy equivalence $f\colon M \to X$. We described in Section 22.1.1 how (M, f) determines a degree one normal map (M, f, ξ, b) to X. Apply Wall realisation to this and a representative $(H(P), P, G)$ of a given class in $L_5^s(\mathbb{Z}[\pi_1(X)])$ to obtain some (M'', f'', ξ'', b''). The result of the action of $(H(P), P, G)$ on (M, f) is defined to be $(M'', f'') \in \mathcal{S}^s(X)$.

This action is independent of the choice of realising 5-manifold. Indeed, suppose (W, F, Ξ, B) and (W', F', Ξ', B') each realise $x \in L_5^s(\mathbb{Z}[\pi_1(X)])$ and are cobordisms from (M, f, ξ, b) to (N, g, θ, c) and (N', g', θ', c') respectively. Construct $(V, G, \Theta, C) := -(W, F, \Xi, B) \cup_{(M,f,\xi,b)} (W', F', \Xi', B;)$, a cobordism from N to N'. The 5-dimensional surgery obstruction of (V, G, Θ, C) vanishes, since it is the difference of the (equal) surgery obstructions of (W, F, Ξ, B) and (W', F', Ξ', B'). By the main theorem of odd dimensional surgery [**Wal99**], (V, G, Θ, C) is bordant relative to the boundary to a simple homotopy equivalence, proving that (N, g) and (N', g') are equal in the structure set $\mathcal{S}^s(X)$. A similar argument as above shows that equivalent formations induce the same action on $\mathcal{S}^s(X)$ and thus we have a well defined action of $L_5^s(\mathbb{Z}[\pi_1(X)])$ on the structure set $\mathcal{S}^s(X)$.

Now we are ready to prove exactness at the structure set. This is stronger than exactness of pointed sets. Precisely, it means that the orbits of the action coincide with the preimages of singleton sets in $\mathcal{N}(X)$.

First observe that (M, f) and (M'', f'') determine the same class in $\mathcal{N}(X)$ because the elements (M, f, ξ, b) and (M'', f'', ξ'', b'') are degree one normally bordant via (W, F, Ξ, B), by construction. We also need to argue that normally bordant homotopy equivalences are related by the action of $L_5^s(\mathbb{Z}[\pi_1(X)])$. Let (M, f) and (N, g) represent elements of $\mathcal{S}^s(X)$ and suppose that (W, F, Ξ, B) is a degree one normal bordism over X between the associated degree one normal maps (M, f, ξ, b) and (N, g, θ, c). Let $\sigma(W, F, \Xi, B) \in L_5^s(\mathbb{Z}[\pi_1(X)])$ be the odd dimensional surgery obstruction of the cobordism. Realise $\sigma(W, F, \Xi, B)$ by a cobordism (W', F', Ξ', B') from (M, f, ξ, b) to some (M', f', ξ', b') using Wall realisation. We claim that (N, g) and (M', f') are equal in the structure set, so that $[(N, g)]$ is in fact obtained from

$[(M,f)]$ by the action of $L_5^s(\mathbb{Z}[\pi_1(X)])$. To see this claim, construct

$$(V,G,\Theta,C) := -(W,F,\Xi,B) \cup_{(M,f,\xi,b)} (W',F',\Xi',B;),$$

a cobordism from N to M'. The 5-dimensional surgery obstruction of (V,G,Θ,C) vanishes, since it is the difference of the (equal) surgery obstructions of (W,F,Ξ,B) and (W',F',Ξ',B'), so (V,G,Θ,C) is bordant relative to the boundary to a simple homotopy equivalence, proving that (M',f') and (N,g) are equal in the structure set $\mathcal{S}^s(X)$.

In fact, by the category preserving 5-dimensional s-cobordism theorem (Section 21.5), N and M' are homeomorphic, but we do not need this.

22.2. The surgery sequence for manifolds with boundary

In many situations, such as when addressing questions about sliceness of knots and links, we will require a generalisation of the material from the previous section to topological manifolds with nonempty boundary, as we will soon see in Chapter 23. We describe now the necessary modifications. For manifolds with boundary, the surgery sequence has the form

$$L_5^s(\mathbb{Z}[\pi_1(X)]) \to \mathcal{S}^s(X,h) \to \mathcal{N}(X,h) \xrightarrow{\sigma} L_4^s(\mathbb{Z}[\pi_1(X)]).$$

We will define the terms, including h, below. Firstly, a 4-dimensional *Poincaré pair* $(X,\partial X)$ is a finite CW complex X together with a subcomplex ∂X, an orientation character $w\colon \pi_1(X) \to \mathbb{Z}/2$, and a *fundamental class* $[X] \in H_4(X,\partial X;\mathbb{Z}^w)$, satisfying the following. Cap product induces simple chain homotopy equivalences

$$-\cap [X]\colon C^{4-*}(X,\partial X;\mathbb{Z}[\pi_1(X)]^w) \xrightarrow{\simeq} C_*(X;\mathbb{Z}[\pi_1(X)]),$$

and

$$-\cap [X]\colon C^{4-*}(X;\mathbb{Z}[\pi_1(X)]^w) \xrightarrow{\simeq} C_*(X,\partial X;\mathbb{Z}[\pi_1(X)]),$$

each connected component $\partial_i X$ of ∂X inherits the structure of a 3-dimensional Poincaré complex with respect to the orientation character induced by w, and a fundamental class $[\partial_i X]$ given by the image of $[X]$ under the homomorphism $H_4(X,\partial X;\mathbb{Z}) \to H_3(\partial X;\mathbb{Z}) \cong \bigoplus_i H_3(\partial_i X;\mathbb{Z})$.

For any compact, topological 4-manifold M, oriented if orientable, the pair $(M,\partial M)$ can be given the structure of a Poincaré pair in a unique way up to simple homotopy equivalence [**KS77**, Theorem 5.13, Essay III].

Fix a 4-dimensional Poincaré pair $(X,\partial X)$. Working 'relative to the boundary' means that we fix a 3-dimensional topological manifold $N = \bigsqcup_i N_i$ with the same number of connected components as ∂X, and a map $h\colon N \to \partial X$ that restricts to a degree one normal map on each connected component. We moreover insist that h induces a simple chain homotopy equivalence $h_*\colon C_*(N_i;\mathbb{Z}[\pi_1(X)]) \to C_*(\partial_i X;\mathbb{Z}[\pi_1(X)])$ for each connected component. The assumption that h is a normal map is required to define a relative normal map set. The $\mathbb{Z}[\pi_1(X)]$ coefficient chain homotopy equivalence is required so that the intersection form on the surgery kernel is nonsingular and the surgery obstruction map σ well defined.

Remark 22.2. For $(X,\partial X)$ a Poincaré pair with $\pi_1(X)$ free, it is sufficient to require h to be a degree one normal map on ∂N inducing a $\mathbb{Z}[\pi_1(X)]$-homology equivalence. This is because in this case the condition that h induce a simple chain homotopy equivalence may be easily verified. The Whitehead group of free groups is trivial by [**BHS64**, Theorem 2], so any chain homotopy equivalence is simple. Next, a chain map of chain complexes of finitely generated projective modules over any ring, bounded above or below, is a chain homotopy equivalence if and only if it is a homology equivalence (see [**Wei94**, Section 10.4]).

The *relative structure set* $\mathcal{S}^s(X,h)$ consists of equivalence classes of pairs (M,f), where M is a compact, topological 4-manifold with boundary N, and $f\colon M \to X$ is a simple homotopy equivalence such that $h = f|_{\partial M}\colon \partial M \to \partial X$. The equivalence relation is defined by setting $(M,f) \sim (M',f')$ when there exists a cobordism $F\colon W \to (X,\partial X)\times[0,1]$ with boundary $\partial(W,F) = -(M,f)\cup_{N\times\{0\}}((N,h)\times[0,1])\cup_{N\times\{1\}}(M',f')$ such that F is a simple homotopy equivalence. We say that (M,f) and (M',f') are *cobordant* over X. Such a cobordism W is in particular an s-cobordism relative to the boundary N.

Since $h\colon N \to \partial X$ is a degree one normal map, this includes the information of a choice of lift of the Spivak normal fibration to $BTOP$ for each connected component $\partial_i X$ of ∂X. The set of *relative normal maps* $\mathcal{N}(X,h)$ is the set of degree one normal bordism classes of *degree one normal maps over* X *relative to* h. These are, by definition, quadruples (M,f,ξ,b), where M is a compact, topological 4-manifold with boundary N, the map $f\colon M\to X$ has degree one and restricts to h on the boundary, and (ξ,b) is stable normal data, covering f and restricting to the given lifts of the Spivak normal fibration on ∂X.

With these modifications, the arguments, definitions, and descriptions of Section 22.1 apply to the interior of M and the exact sequence given above may be constructed similarly.

22.3. Classification of closed, simply connected 4-manifolds

In the rest of this chapter we explain the application of the surgery exact sequence to the homeomorphism classification of closed, simply connected, topological 4-manifolds, stated below.

THEOREM 22.3 (Classification of closed, simply connected, topological 4-manifolds, [**Fre82a**, Theorem 1.5; **FQ90**, Chapter 10]). *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, simply connected, topological, (spin), oriented 4-manifold, unique up to homeomorphism, whose second homology is isomorphic to F and whose intersection form is isometric to 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, simply connected, topological, (non-spin), oriented 4-manifolds with intersection form isometric to θ, one of which is stably smoothable and one of which is not.

Let M and M' be two closed, 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 (in other words, their Kirby-Siebenmann invariants coincide). 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 corresponds to a closed, simply connected, topological 4-manifold, unique up to homeomorphism, which realises it as its intersection form. Such a manifold is stably smoothable if and only if the signature is divisible by 16. For symmetric, integral matrices with determinant ±1 which are odd instead, there are, up to homeomorphism, two closed, simply connected, topological 4-manifolds, exactly one of which is stably smoothable. Moreover, any isometry of forms is realised by a homeomorphism. Note that homeomorphism preserves stable smoothability.

Quinn [**Qui86**] showed moveover that such homeomorphisms are unique up to isotopy, but we will not discuss this here, referring the interested reader to Quinn's paper, and to [**CH90**] where some errors from [**Qui86**] are corrected.

The classification that we will outline proceeds as follows. First, Whitehead [**Whi49**] and Milnor [**Mil58**] showed the following.

THEOREM 22.4 (Milnor-Whitehead classification). *Homotopy equivalence classes of 4-dimensional, oriented, simply connected Poincaré complexes X are in one to one correspondence with nonsingular, symmetric, bilinear forms over $\mathbb{Z}$, with the correspondence given by taking the intersection form $\lambda_X\colon H_2(X;\mathbb{Z}) \times H_2(X;\mathbb{Z}) \to \mathbb{Z}$.*

Moreover, any isometry of nonsingular, symmetric forms is induced by a homotopy equivalence of the corresponding Poincaré complexes.

For the last statement in the theorem above, see [**Whi49**, Theorems 3 and 4]. In the asserted correspondence, a given nonsingular, symmetric, bilinear form is realised by attaching a 4-cell to a wedge $\bigvee_r S^2$ by a suitable map in $\pi_3(\bigvee_r S^2)$. For the desired classification of manifolds, we first compute the structure set for simply connected 4-dimensional Poincaré complexes. We have seen that when $\pi_1(X)$ is good, we have the following exact sequence of pointed sets:

$$L_5^s(\mathbb{Z}[\pi_1(X)]) \to \mathcal{S}^s(X) \to \mathcal{N}(X) \xrightarrow{\sigma} L_4^s(\mathbb{Z}[\pi_1(X)]).$$

Here the first map is really an action, whose orbits coincide with preimages in $\mathcal{S}^s(X)$ of singletons in $\mathcal{N}(X)$. We are currently interested in the case where $\pi_1(X)$ is trivial. First, note that the trivial group is good (see Chapter 19). Moreover, the Whitehead group of the trivial group is itself trivial, and so all homotopy equivalences are simple with respect to any choice of chain basis. Thus $\mathcal{S}(X) = \mathcal{S}^s(X)$, where the former was defined in Chapter 1 and consists of pairs $(M, f\colon M \to X)$ where f is a homotopy equivalence, modulo h-cobordism. The odd dimensional simply connected L-groups are also trivial [**Wal99**, Theorem 13A.1], so $L_5^s(\mathbb{Z}) = 0$. Additionally, by [**MH73**, II.4.4, II.5.1, and IV.2.7] we have $L_4^s(\mathbb{Z}[\{1\}]) \cong 8\mathbb{Z}$, where the map $\sigma\colon \mathcal{N}(X) \to L_4^s(\mathbb{Z})$ sends $(M, f, \xi, b) \in \mathcal{N}(X)$ to the signature difference $\operatorname{sign}(M) - \operatorname{sign}(X) \in 8\mathbb{Z}$ [**Ran02**, Example 11.77]. The surgery sequence for a simply connected 4-dimensional Poincaré complex thus reduces to

$$0 \to \mathcal{S}(X) \to \mathcal{N}(X) \xrightarrow{\sigma} 8\mathbb{Z}.$$

Recall also the fully topological h-cobordism theorem (Section 21.5) which states that any 5-dimensional, simply connected, topological h-cobordism between closed 4-manifolds is homeomorphic to a product.

The surgery sequence is also a mnemonic for the fact that every degree one normal map $(M, f, \xi, b) \in \mathcal{N}(X)$ with $\sigma((M, f, \xi, b)) = 0 \in L_4^s(\mathbb{Z}[\pi_1(X)])$ is normally bordant to an element of $\mathcal{S}^s(X)$. This holds even when we do not know the structure set is nonempty, so can be used to deduce that. Indeed, we will use this fact to show that $\mathcal{S}(X)$ is nonempty for every simply connected, oriented, 4-dimensional Poincaré complex X. In other words, that there exists a closed, topological 4-manifold M homotopy equivalent to X. Then we will compute the orbit space $\mathcal{S}(X)/\operatorname{hAut}(X)$ under the action of orientation preserving self-homotopy equivalences of X on $\mathcal{S}(X)$. This set of orbits agrees with the set of homeomorphism classes of closed, oriented topological manifolds M within the oriented homotopy equivalence class of X, as explained in Section 22.1. In order to obtain the classification above, we will show that this quotient contains one element when the intersection form is even and contains two elements when the intersection form is odd.

22.3.1. Existence of a 4-manifold. Fix a simply connected 4-dimensional Poincaré complex X. We first need to find a closed, topological 4-manifold M with a degree one normal map to X, that is a degree one map $f\colon M \to X$ covered by a stable bundle map

$$\begin{tikzcd} \nu_M \arrow[r, "b"] \arrow[d] & \xi \arrow[d] \\ M \arrow[r, "f"] & X \end{tikzcd}$$

where $\xi \to X$ is a lift of the Spivak normal fibration of X. For simply connected 4-dimensional Poincaré complexes, this can always be constructed by Remark 22.1. By the exactness of the surgery sequence, the surgery obstruction $\sigma(M, f, \xi, b)$ vanishes if and only if there is a normal bordism to a homotopy equivalence. As noted above, the surgery obstruction vanishes if and only if $\operatorname{sign}(M) - \operatorname{sign}(X) = 0$, and indeed we can arrange that M and X have equal signatures as follows. It is an algebraic fact that the signature of an even, symmetric, integral, nonsingular, bilinear form is divisible by eight (see [**MH73**, §2, Theorem 5.1]). The next two propositions will show the existence of the E_8*-manifold*, a closed, simply connected, topological 4-manifold whose intersection form is the E_8 form. By taking the connected sum of M with copies of the E_8-manifold with either orientation, we can change the signature of M by multiples of eight until it equals the signature of X.

Once the surgery obstruction vanishes, the exactness of the surgery sequence implies that there exists a closed, topological manifold M' homotopy equivalent to X. The discussion above implies that M' is obtained from M by using the category preserving sphere embedding theorem and then performing surgery on the framed, embedded spheres thus produced, which represent half a basis for the surgery kernel.

Remark 22.5. The result of Land [**Lan17**, Theorem 3.3] mentioned in Remark 22.1 in fact shows that every oriented 4-dimensional Poincaré complex X admits a degree one normal map from a smooth, closed 4-manifold M. However, once we take the connected sum of M with copies of the E_8-manifold, we lose smoothness. Consequently, the category losing version of the sphere embedding theorem is not sufficient.

The next two propositions complete the proof that the structure set is nonempty for every 4-dimensional simply connected Poincaré complex. In other words, every symmetric, nonsingular, bilinear form over $\mathbb{Z}$ is realised as the intersection form of a closed, simply connected, topological 4-manifold. See Section 21.3.2 for more discussion as well as a sketch of an alternative proof.

Proposition 22.6. *Every 3-dimensional integral homology sphere bounds a compact, contractible, topological 4-manifold.*

Proof. Let Σ be a 3-dimensional integral homology sphere. Consider the (smooth) 4-manifold $\Sigma \times [0,1]$. We can find a generating set for $\pi_1(\Sigma)$, pushed into the interior of $\Sigma \times [0,1]$, bounding a mutually disjoint collection of surfaces $\{F_i\}$, since $H_1(\Sigma) = 0$. Perform surgery on this generating set with respect to the framing induced by the surfaces, removing copies of $S^1 \times D^3$ and glueing in copies of $D^2 \times S^2$. This changes $\Sigma \times [0,1]$ into a (smooth) simply connected 4-manifold V', which still has $\partial V' = \Sigma \sqcup -\Sigma$, such that $H_2(V')$ is generated by $\{F_i\}$ capped off by the surgery discs $\{D_i\}$, along with geometrically transverse spheres $\{g_i\}$ corresponding to $\{*\} \times S^2 \subseteq D^2 \times S^2$.

The generators of the fundamental group of $\{F_i \cup D_i\}$ bound immersed discs in V', since $\pi_1(V')$ is trivial. Use these immersed discs, with the correct framing, to contract the surfaces to framed immersed spheres $\{f_i\}$. The spheres are framed since the surfaces $\{F_i \cup D_i\}$ were and we contracted using framed, immersed discs. We also note that $\lambda(f_i, f_j) = \mu(f_i) = 0$ for all i, j. Moreover, the set $\{g_i\}$ is algebraically transverse to the spheres $\{f_i\}$.

Now we apply the sphere embedding theorem (Section 20.3), in the simply connected (smooth) manifold V', to obtain embedded spheres $\{\overline{f}_i\}$, regularly homotopic to $\{f_i\}$ and equipped with geometrically transverse spheres $\{\overline{g}_i\}$. Perform surgery on the embedded spheres $\{\overline{f}_i\}$ to obtain a topological 4-manifold V which still has $\partial V = \Sigma \sqcup -\Sigma$. Note that this second surgery step does not introduce nontrivial elements of $\pi_1(V)$ due to the geometrically transverse spheres $\{\overline{g}_i\}$. We see that $\pi_1(V)$ and $H_2(V)$ are both trivial.

Following the proof of [**FQ90**, Theorem 9.3C], stack together infinitely many copies of V, and construct the endpoint compactification W. Lastly, appeal to [**FQ90**, Theorem 9.3A] to see that W is indeed a manifold at the extra point added in to compactify. This final step uses work of Quinn [**Qui82b**]. □

A different proof of the Proposition 22.6, using the topological surgery sequence for pairs, is given in [**FQ90**, Section 11.4]. However note that that proof implicitly assumes the existence of the E_8-manifold, so we cannot use this proof to obtain the next proposition. More discussion on these kinds of issues can be found in Chapter 21.

PROPOSITION 22.7. *There exists a closed, simply connected, topological* 4*-manifold whose intersection form is isometric to the form* $\lambda\colon \mathbb{Z}^8 \times \mathbb{Z}^8 \to \mathbb{Z}$ *represented by the* E_8 *matrix*

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

PROOF. Plumb together eight copies of the D^2 bundle over S^2 with Euler number two, that is the unit disc bundle of the tangent bundle of S^2, according to the E_8 matrix. This results in a compact, smooth 4-manifold X with boundary an integral homology sphere Σ; indeed Σ is the Poincaré homology sphere. By Proposition 22.6, Σ bounds a contractible, topological 4-manifold W. Cap off the manifold X with W to obtain a closed, topological manifold, the E_8-manifold. □

22.3.2. Size of the structure set. Now that we know the structure set of X is nonempty, we want to see how many elements it has. For this we again investigate the surgery sequence. The set of normal invariants $\mathcal{N}(X)$ is in bijective correspondence with

$$[X, G/TOP] \cong H^2(X; \mathbb{Z}/2) \oplus H^4(X; \mathbb{Z}),$$

as follows. The spaces BG and $BTOP$ were defined in Section 22.1.4. The space G/TOP is by definition the homotopy fibre of the forgetful map $BTOP \to BG$. The identification of normal bordism classes of degree one normal maps with $[X, G/TOP]$ uses the Pontryagin-Thom construction [**Lüc02**, Chapter 3], and in particular uses topological transversality (specifically, the case of map transversality where the inverse image has dimension four (see Section 21.4.9)). As mentioned in Section 21.5.1, there is a 5-connected map $G/TOP \to K(\mathbb{Z}/2, 2) \times$

$K(\mathbb{Z}, 4)$ [**KT01**, p. 397] whence the identification $[X, G/TOP] \cong H^2(X;\mathbb{Z}/2) \oplus H^4(X, \mathbb{Z})$. Now

$$H^2(X;\mathbb{Z}/2) \oplus H^4(X;\mathbb{Z}) \cong H_2(X;\mathbb{Z}/2) \oplus H_0(X;\mathbb{Z}) \cong H_2(X;\mathbb{Z}/2) \oplus \mathbb{Z}.$$

Moreover, the fibre under the surgery obstruction map of each element of $L_4^s(\mathbb{Z})$ is $H_2(X;\mathbb{Z}/2)$. To see this, use that the map $\mathcal{N}(X) \to L_4^s(\mathbb{Z}) \cong 8\mathbb{Z}$ sends a degree one normal map $f\colon M \to X$ to $\sigma(M) - \sigma(X)$ and that the bijection $\mathcal{N}(X) \to H_2(X;\mathbb{Z}/2) \oplus \mathbb{Z}$ sends $f\colon M \to X$ to $(\kappa(f), (\sigma(M) - \sigma(X))/8)$, where κ is a codimension two Arf-Kervaire invariant [**Dav05**, Proposition 3.6]. Part of [**Dav05**, Proposition 3.6] also states that the induced map $H_2(X;\mathbb{Z}/2) \to L_4(\mathbb{Z}[\pi_1(X)])$ factors through $H_2(B\pi_1(X);\mathbb{Z}/2)$, which in the case of trivial $\pi_1(X)$ is of course also trivial. Thus the fibre of the surgery obstruction map is $H_2(X;\mathbb{Z}/2)$ as asserted above.

In other words, the surgery sequence can be written as

$$0 \to \mathcal{S}(X) \to H_2(X;\mathbb{Z}/2) \oplus \mathbb{Z} \xrightarrow{(0,\times 8)} 8\mathbb{Z}.$$

It follows that the structure set is

$$\mathcal{S}(X) \cong \ker(\sigma) \cong H_2(X;\mathbb{Z}/2).$$

Finally, we appeal to work of Cochran and Habegger [**CH90**, Theorem 5.1], who showed that the action of the orientation preserving self-homotopy equivalences of X on $\mathcal{S}(X)$ is transitive on $\ker(w_2(X)\colon H_2(X;\mathbb{Z}/2) \to \mathbb{Z}/2)$. It follows that, when the form θ is even, since $w_2(X) = 0$, the action is transitive. When θ is odd, $\ker(w_2(X))$ is an index two subgroup and so the orientation preserving self-homotopy equivalences have two orbits. Hence when θ is even there is a unique closed, oriented, topological 4-manifold homotopy equivalent to the Poincaré complex X, up to homeomorphism, and when θ is odd there are two closed, oriented, topological 4-manifolds homotopy equivalent to X, up to homeomorphism.

It remains to explain which elements are stably smoothable. For this we use the Kirby-Siebenmann invariant, and appeal to the fact that a closed, topological 4-manifold M is stably smoothable if and only if the Kirby-Siebenmann invariant $\operatorname{ks}(M) \in \mathbb{Z}/2$ vanishes [**FQ90**, Theorem 8.6; **FNOP19**, Theorem 8.6]. By definition, the Kirby-Siebenmann invariant is the (complete) obstruction for the stable tangent microbundle of M to admit a lift to BPL. 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 [**FQ90**, Chapter 8].

For closed, simply connected, topological 4-manifolds with even intersection form, the Kirby-Siebenmann invariant is congruent mod 2 to $\sigma(M)/8$ [**FQ90**, Section 10.2B; **FNOP19**, Theorem 8.2(4)]. Now for the case that the intersection form is odd, we proceed by constructing, from a topological manifold N with that intersection form and $\operatorname{ks}(N) = 0$, another topological manifold $*N$ with the same intersection form and opposite Kirby-Siebenmann invariant. Thus N and $*N$ are homotopy equivalent but not homeomorphic, the manifold N is stably smoothable, and the manifold $*N$ is not.

To construct $*N$ we first need to construct $*\mathbb{CP}^2$. This was explained in Chapter 1, but for convenience we repeat the construction here. Attach a $+1$-framed 2-handle to D^4 along a knot K in $S^3 = \partial D^4$ with $\operatorname{Arf}(K) = 1$. 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 Dehn surgery on S^3 along K. Cap off this homology sphere with the contractible, topological 4-manifold C with boundary Σ promised by Proposition 22.6 and call the resulting topological manifold N'. To see that N' is homeomorphic to $*\mathbb{CP}^2$, it suffices to show that $\operatorname{ks}(N') = 1$. For this, observe that $\operatorname{ks}(N') = \operatorname{ks}(C)$ since the Kirby-Siebenmann invariant is additive for 4-manifolds glued along their boundary [**FQ90**, Section 10.2B; **FNOP19**, Theorem 8.2(5] and $X_1(K)$ is smooth, modulo smoothing corners. Since C is contractible, C is a topological spin manifold. By [**FQ90**, Section 10.2B; **GA70**], we see that $\mathrm{ks}(C) = \mu(\Sigma) = \mathrm{Arf}(K) = 1$ where $\mu(\Sigma)$ is the Rochlin invariant of Σ. To see independence of the choice of knot K, note that the resulting manifolds will have equal Kirby-Siebenmann invariant and are thus homeomorphic by our previous work.

Now that we have constructed $*\mathbb{CP}^2$ we proceed to construct $*N$ for a closed, simply connected, toplogical 4-manifold N with odd intersection form θ. Consider $N\#*\mathbb{CP}^2$. This has intersection form $\theta \oplus \langle 1\rangle$. The $\langle 1\rangle$ summand can be represented by an embedded sphere whose normal bundle has Euler number 1; this uses a version of the sphere embedding theorem with unframed transverse spheres and the fact that $w_2(N\# * \mathbb{CP}^2)$ is nonzero on the orthogonal complement θ of $\langle 1\rangle$ [**FQ90**, Section 10.3] (see also [**Sto94**]). Thus we can write

$$N\# * \mathbb{CP}^2 \cong *N\#\mathbb{CP}^2$$

for some closed, topological 4-manifold $*N$ with intersection form isometric to θ. Since $\mathrm{ks}(*\mathbb{CP}^2) = 1$ and $\mathrm{ks}(\mathbb{CP}^2) = 0$, the fact that ks is additive with respect to connected sum [**FQ90**, Section 10.2B; **FNOP19**, Theorem 8.2(3)] implies that $\mathrm{ks}(*N) = 1$. This completes the construction of a closed, simply connected, topological 4-manifold $*N$ that is homotopy equivalent to N but not homeomorphic, with opposite Kirby-Siebenmann invariant. These two manifolds represent the two distinct orbits in the structure set of N. As stated above, the manifold N is stably smoothable and the manifold $*N$ is not.

22.3.3. Realising isometries by homeomorphisms. Suppose we have two closed, simply connected, topological 4-manifolds M and N together with an isomorphism $\theta\colon H_2(M;\mathbb{Z}) \xrightarrow{\cong} H_2(N;\mathbb{Z})$ inducing an isometry between the intersection forms λ_M and λ_N. Suppose that the Kirby-Siebenmann invariants of M and N coincide. We want to show that there is a homeomorphism inducing θ. By the Milnor-Whitehead classification (Theorem 22.4) there is a homotopy equivalence $f\colon M \simeq N$ inducing the given map θ on second homology. By the previous section, we can modify f by a self-homotopy equivalence of N such that $(M, f\colon M \to N)$ and $(N, \mathrm{Id}\colon N \to N)$ become equal in the structure set $\mathcal{S}(N)$, if necessary. By the work of Cochran and Habegger [**CH90**, Theorem 5.1], this can be done by a self-homotopy equivalence of N that induces the identity on second homology. Then by the fact that $f\colon M \to N$ and $\mathrm{Id}\colon N \to N$ are equal in $\mathcal{S}(N)$, and by the category preserving h-cobordism theorem (Section 21.5), there is a homeomorphism $g\colon M \xrightarrow{\cong} N$ such that the diagram

$$\begin{tikzcd} M \arrow[rr, "\cong", "g"'] \arrow[dr, "\simeq"', "f"'] & & N \arrow[dl, "\mathrm{Id}"] \\ & N & \end{tikzcd}$$

commutes up to homotopy. In particular f is homotopic to a homeomorphism, as required. Since g and f are homotopic, they induce the same maps on second homology.

22.3.4. Other homeomorphism classifications of 4-manifolds. The strategy used above follows the high dimensional surgery programme for classifying manifolds: calculate the structure set using the surgery exact sequence, then take the quotient by simple self-homotopy equivalences. Both of the stages are more difficult when the fundamental group is nontrivial. Calculating the collection of simple self-homotopy equivalences of a non-simply connected 4-dimensional Poincaré complex is considered to be a particularly hard problem in general. We briefly describe two alternative classification strategies and how they have been successfully applied to certain good fundamental groups.

One alternative approach was given in [**Fre82a**] and is also described in [**FQ90**, Chapter 10]. In this approach one analyses the following question. Given a closed, simply connected, topological 4-manifold M with intersection form θ, and a direct sum decomposition of θ as a form, when is that decomposition induced by a connected sum of simply connected 4-manifolds $M = M'\#M''$? This question has a precise existence and uniqueness answer. The idea is then to use the algebraic classification of nonsingular integral intersection forms, together with explicit constructions of the E_8-manifold and $*\mathbb{C}P^2$ to first build 4-manifolds corresponding to all nonsingular integral intersection forms and then to use the connected sum theorem to analyse the uniqueness of the constructions.

A variant of the approach just described has been applied when the fundamental group is $\mathbb{Z}$. The latter group is good by Theorem 19.4. Here a strong existence and uniqueness result similar to Theorem 22.3 holds [**FQ90**, Theorem 10.7A]. In this case, the relevant question is: given a closed, topological 4-manifold M with fundamental group $\mathbb{Z}$, and $\mathbb{Z}[\mathbb{Z}]$ intersection form θ, together with a direct sum decomposition of θ as a $\mathbb{Z}[\mathbb{Z}]$-form, is this decomposition induced by a fibre sum $M = M'\#_{S^1}M''$? Here, M' and M'' have fundamental group $\mathbb{Z}$ and the fibre sum is taken over a generator. The answer to this question and the remainder of this programme are described in [**FQ90**, Chapter 10].

A second alternative approach uses Kreck's modified surgery theory [**Kre99**]. The papers [**HK88**, **HK93**] use this theory to show that a closed, topological 4-manifold M with finite cyclic fundamental group is determined up to homeomorphism by the isomorphism class of the integral intersection form on $H_2(M;\mathbb{Z})/\text{Torsion}$, together with the Kirby-Siebenmann invariant and its *w_2-type*. Here two 4-manifolds have the same w_2-type if they are either both spin, both have non-spin universal covers, or both are non-spin but with spin universal cover. Partial results towards the realisation of such closed 4-manifolds are also given. Recall that finite groups are good by Theorem 19.2.

In the case of non-cyclic fundamental groups, closed 4-manifolds with fundamental group a solvable Baumslag-Solitar group $B(k) := \langle a, b \mid aba^{-1}b^{-k}\rangle$ are classified in terms of their equivariant intersection forms and the Kirby-Siebenmann invariant in [**HKT09**]. There is also a corresponding realisation result [**HKT09**, Theorem B]. Recall that solvable groups are good (Example 19.6). Partial results towards a classification for 4-manifolds with good fundamental groups with cohomological dimension 3 were given in [**HH19**].

Closed, nonorientable 4-manifolds with fundamental group $\mathbb{Z}/2$ were classified in [**HKT94**], where a complete list of such manifolds was given. The existence and uniqueness of closed, nonorientable 4-manifolds with infinite cyclic fundamental group in terms of the equivariant intersection form and the Kirby-Siebenmann invariant was given in [**Wan95**].

Some results have also been obtained using the classical surgery strategy. A formidable calculation of surgery obstruction groups for the infinite dihedral group $\mathbb{Z}/2 * \mathbb{Z}/2$ leads to a computation of the structure set $\mathcal{S}^s(\mathbb{RP}^4\#\mathbb{RP}^4)$ of closed, topological 4-manifolds homotopy equivalent to $\mathbb{RP}^4\#\mathbb{RP}^4$ in [**BDK07**] (see also [**JK06**]). Recall that $\mathbb{Z}/2 * \mathbb{Z}/2$ is a good group since it can be expressed as an extension of good groups (Proposition 19.5(3)).

One can also classify some aspherical 4-manifolds up to homeomorphism, when the fundamental group is both good and satisfies the Borel conjecture (see, for example, [**KL04**]). The simplest case is the rigidity of the 4-torus.

THEOREM 22.8. *Let M be a closed, topological 4-manifold homotopy equivalent to the 4-torus T^4. Then M is homeomorphic to T^4.*

Finally, compact, simply connected, topological 4-manifolds with fixed connected nonempty boundaries have been classified using the sphere embedding theorem [**Boy86**, **Boy93**, **Vog82**, **Sto93**] in terms of the intersection form and the Kirby-Siebenmann invariant. Here it does not suffice just for the intersection forms to be isometric, rather they must be isometric presentation forms of the linking form on the torsion subgroup of the first homology of the boundary 3-manifold.

CHAPTER 23

Open problems

Min Hoon Kim, Patrick Orson, JungHwan Park, and Arunima Ray

We finish Part III by describing some of the major open problems related to the disc embedding theorem. The statement of the disc embedding theorem without the good group hypothesis and within a topological ambient manifold is called the *disc embedding conjecture*. As explained in Section 21.5, while the proof in this book requires a smooth ambient manifold, a category preserving disc embedding theorem, assuming good fundamental group, can be established using the same proof. So, if all groups were good, this would establish the disc embedding conjecture, which in turn would imply the s-cobordism theorem and the exactness of the surgery sequence with no restriction on fundamental groups, via the proofs given in Chapters 20 and 22. Thus the disc embedding conjecture is of central importance in the field of topological 4-manifolds. Several reformulations and implications of the disc embedding conjecture have been found, and we describe these in this chapter. They are summarised in Figure 23.1.

The structure of this chapter is as follows. First we will explain the top row of Figure 23.1, consisting of the equivalence of the disc embedding conjecture with Conjectures G and $\mathcal{L}_1$. Then we will introduce the *surgery conjecture* (Conjecture 23.8). This has its own set of conjectures equivalent to it, shown in Figure 23.3. We present these equivalences, then at the end of the chapter we return to Figure 23.1 and explain the lower two boxes.

As in the previous chapter, we will implicitly use the immersion lemma, topological transversality, and the existence of normal bundles for locally flat submanifolds of a topological manifold in order to work solely in the topological category.

23.1. The disc embedding conjecture

CONJECTURE 23.1 (Disc embedding conjecture). *Let M be a connected, topological 4-manifold with nonempty boundary. Let*

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

be an immersed collection of discs in M with pairwise disjoint, embedded boundaries. Suppose there is an immersed collection of framed dual 2-spheres

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

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

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

with an immersed collection of framed geometrically dual spheres

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

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

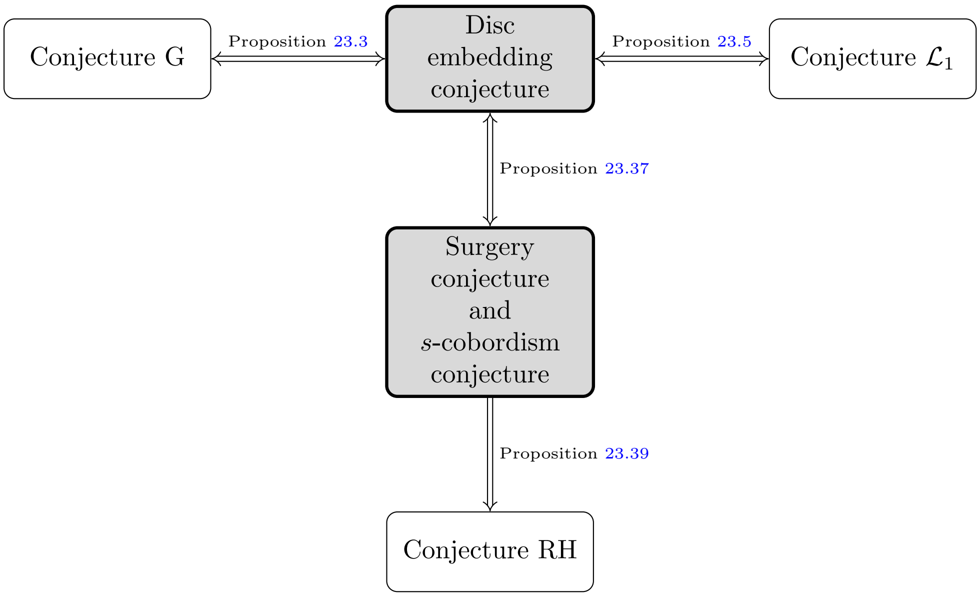


Figure 23.1. The disc embedding conjecture in relation to other conjectures.

Capped gropes feature prominently in the proof of the disc embedding theorem presented in this book. Unsurprisingly, there is a direct connection between properties of capped gropes and the disc embedding conjecture. This is established in the following conjecture and equivalence.

Conjecture 23.2 (Conjecture G). *Every disc-like capped grope G^c of height 1.5 contains a flat embedded disc whose framed boundary coincides with the attaching region of G^c.*

Proposition 23.3. *The disc embedding conjecture is equivalent to Conjecture G.*

Proof. Assume the disc embedding conjecture. Given a disc-like capped grope G^c of height 1.5, divide its second and third stages into $(+)$- and $(-)$-sides as usual, such that the $(-)$-side consists of only caps. The idea is to replace the $(-)$-side caps with mutually disjoint, flat, embedded discs. Use two parallel copies of the $(+)$-side, along with an annulus joining them, to form a transverse union-of-spheres-like capped grope of height one for the $(-)$-side caps. The body of this grope is geometrically transverse to the $(-)$-side caps, but its caps may intersect the $(-)$-side caps arbitrarily (see Chapter 17 for more details).

Use the sequential contraction lemma (Lemma 17.7) to produce a mutually disjoint union of immersed spheres (namely, (degenerate) sphere-like gropes of height zero), which are algebraically transverse to the $(-)$-side caps. Note that these would have been geometrically transverse had we applied the cap separation lemma (Lemma 17.4) to separate the $(+)$- and $(-)$-side caps. However, we only need algebraically transverse spheres for the disc embedding conjecture so we did not bother.

Next we observe that the $(-)$-side caps with these transverse spheres satisfy the hypotheses of the disc embedding conjecture in the complement in G^c of the first stage. In particular, the spheres have trivial intersection and self-intersection numbers since they are produced by contraction of mutually disjoint and embeded surfaces. The disc embedding conjecture then allows us to replace the $(-)$-side caps of G^c by mutually disjoint, flat embedded discs with the same framed boundary. Use these to surger the first stage surface of G^c to an embedded, flat disc within G^c

with the desired framed boundary, as desired. This shows that the disc embedding conjecture implies Conjecture G.

For the reverse implication, we need to recall the steps of the proof of the disc embedding theorem. We first replaced the original immersed discs $\{f_i\}$ and algebraically transverse spheres $\{g_i\}$ by geometrically transverse immersed discs $\{f'_i\}$ (with the same framed boundary as $\{f_i\}$) and immersed spheres $\{g'_i\}$, where the intersection and self-intersection points of $\{f'_i\}$ are paired by framed, immersed Whitney discs $\{W_k\}$ with interiors lying in the complement of $\{f'_i\}$ (Proposition 16.1). The discs $\{W_k\}$ have a collection of geometrically transverse capped surfaces $\{\Sigma^c_k\}$ whose bodies are Clifford tori at the double points of $\{f'_i\}$. Next, we replaced the discs $\{W_k\}$ by a union-of-discs-like capped grope $G^c(2)$ of height two, which lies in the complement of $\{f'_i\}$ other than its attaching region, and such that $\{\Sigma^c_k\}$, possibly after a regular homotopy of its caps, is geometrically transverse to the bottom stage surfaces of $G^c(2)$ (Proposition 16.2). In particular, the caps of $\{\Sigma^c_k\}$ do not intersect the caps of $G^c(2)$.

We cannot immediately apply the hypothesis that Conjecture G holds, since $G^c(2)$ might not be a mutually disjoint collection of disc-like capped gropes, as the caps may intersect arbitrarily. Instead, apply grope height raising (Proposition 17.3) to get a grope of height 3, then apply the sequential contraction lemma (Lemma 17.7) to make the caps pairwise disjoint. A subsequent contraction produces $\{G^c(1.5)_k\}$, a collection of pairwise disjoint capped gropes with height 1.5 pairing the intersection points within $\{f'_i\}$ with the correct framed attaching region. Contract the set $\{\Sigma^c_k\}$ to produce geometrically transverse spheres for $\{G^c(1.5)_k\}$. Tube $\{g'_i\}$ into these spheres to obtain the spheres $\{\overline{g}_i\}$ such that the collection $\{G^c(1.5)_k\}$ lies in the exterior of both $\{f'_i\}$ and $\{\overline{g}_i\}$ other than the attaching region, and $\{\overline{g}_i\}$ is geometrically transverse to $\{f'_i\}$. By hypothesis, each element of the set $\{G^c(1.5)_k\}$ contains a flat embedded disc whose framed boundary coincides with the attaching region of the grope. Use these discs, which are mutually disjoint by construction, as Whitney discs, so that the corresponding Whitney move transforms $\{f'_i\}$ into a collection $\{\overline{f}_i\}$ of mutually disjoint, flat embedded discs, with the same framed boundary as the original $\{f_i\}$, and equipped with the geometrically transverse spheres $\{\overline{g}_i\}$, as desired. To see that each $\{\overline{g}_i$ is homotopic to g_i, apply Lemma 17.11. ☐

23.1.1. Standard slices for universal links. The techniques of Kirby calculus from Chapter 13 allow us to rephrase Conjecture G in terms of the sliceness of certain links. We will construct a *universal* family of links, universal in the sense that the disc embedding conjecture is equivalent to these links being slice with a certain *standard* property (Proposition 23.5). An equivalence of this nature was first found by Casson in [**Cas86**], related to the ends of noncompact 4-manifolds. There are other families of universal links, related to the disc embedding and other conjectures, which we describe soon.

Let $\mathcal{L}_1$ denote the class of links obtained from the Hopf link by performing at least two iterations of ramified Bing doubling on a single component, and then performing ramified Whitehead doubling on all of the new components. (Recall that usual Bing and Whitehead doubling are special cases of ramified Bing and ramified Whitehead doubling; see Section 13.5.) The ramification parameters need not be constant across components. All possible choices of clasps for the ramified Whitehead doubling step are allowed. More precisely, given an n-component link $L \in \mathcal{L}_1$, each of the 2^n-1 other links produced by changing the sign of any nontrivial subset of the clasps produced in the ramified Whitehead doubling step also lies in $\mathcal{L}_1$. Note that, in this construction, one of the components of the original Hopf link remains unchanged. For a link $L \in \mathcal{L}_1$, call this component $m(L)$. Observe that if we remove $m(L)$ from L, the resulting link is an unlink.

An example of an element L of $\mathcal{L}_1$ is given in Figure 23.2; the reader might verify that L becomes the unlink after removing the component $m(L)$. Moreover, the ambitious reader should recognise the elements of $\mathcal{L}_1$ as precisely those links arising as Kirby diagrams for disc-like capped gropes, such that the caps are pairwise disjoint, as in Chapter 13. In particular, $m(L)$ is the attaching circle. The requirement of at least two rounds of Bing doubling implies that the corresponding (symmetric) gropes have height at least two. We have the following conjecture and equivalence.

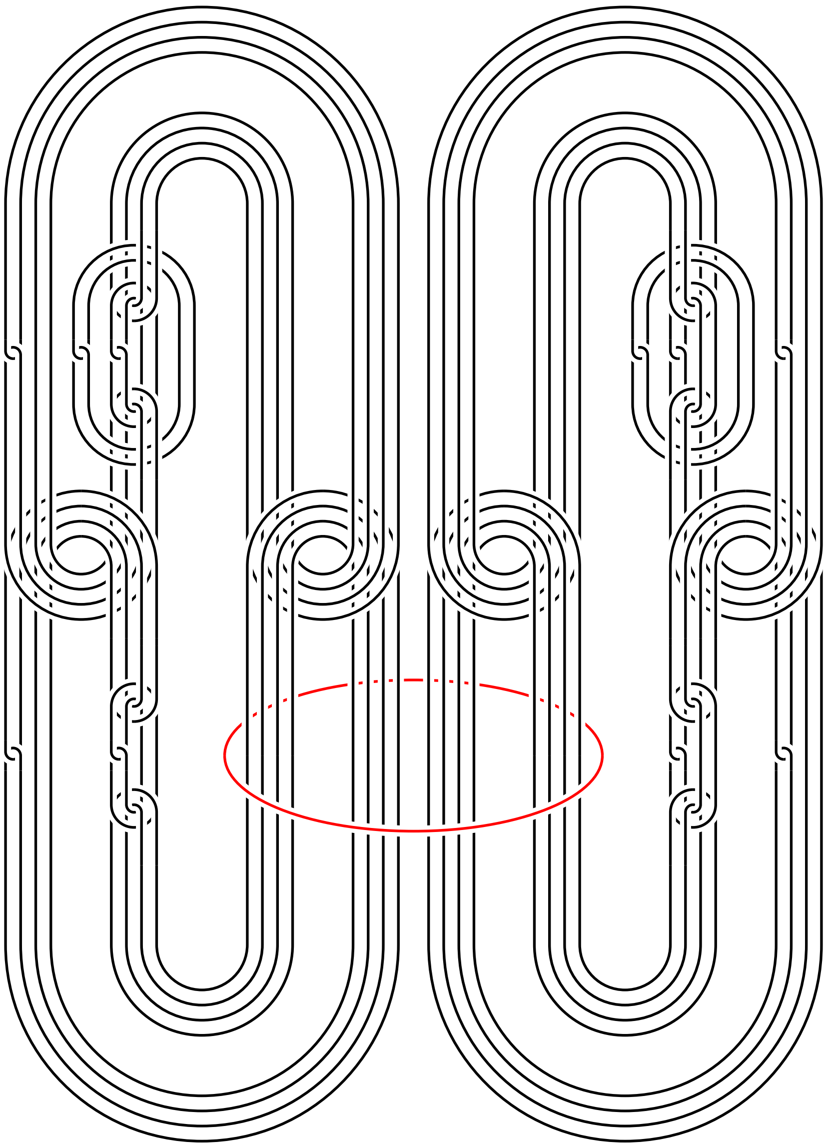


Figure 23.2. An element L of the set $\mathcal{L}_1$. The component $m(L)$ is in red. Compare with Figure 13.23.

Conjecture 23.4 (Conjecture $\mathcal{L}_1$). *Each link L in the family $\mathcal{L}_1$ is topologically slice in D^4, where the slice discs for $L \smallsetminus m(L)$ are standard.*

Here, we say that a collection of discs in D^4 bounded by an unlink is *standard* if they can be isotoped, relative to the boundary, to lie in S^3. Recall that a link in S^3 is said to be *topologically slice* in D^4 if the components of L bound a collection of pairwise disjoint locally flat embedded discs in D^4. The question of whether the links in a universal family are topologically slice is sometimes called a *universal link slicing problem.*

We remark here that by work of Quinn locally flat submanifolds of topological 4-manifolds admit normal bundles (see Section 21.4.8), so in particular any slice link in fact bounds a collection of pairwise disjoint, flat, embedded discs in D^4.

PROPOSITION 23.5. *The disc embedding conjecture is equivalent to Conjecture $\mathcal{L}_1$.*

PROOF. As mentioned above, the links in $\mathcal{L}_1$ are precisely the Kirby diagrams for capped gropes of height two, whose caps are mutually disjoint, where the component $m(L)$ is the attaching circle, and where the remaining components correspond to 1-handles. The key observation is that we obtain a capped grope by removing the standard discs bounded in D^4 by the curves corresponding to the 1-handles, which form an unlink. See Chapter 13 for more details.

Assume Conjecture $\mathcal{L}_1$. Given a disc-like capped grope G^c of height 1.5, apply grope height raising (Proposition 17.3) and sequential contraction (Lemma 17.7) to obtain a grope within G^c, with the same attaching region, height at least two, and pairwise disjoint caps. Then the corresponding Kirby diagram, when considered as a link L (along with the attaching circle $m(L)$), is an element of $\mathcal{L}_1$. The slice disc bounded by $m(L)$ in D^4 is the disc bounded by the attaching circle within G^c. Thus Conjecture $\mathcal{L}_1$ implies that every disc-like capped grope of height 1.5 contains a flat embedded disc whose framed boundary coincides with the attaching region of the capped grope, so Conjecture G holds. According to Proposition 23.3, Conjecture G implies the disc embedding conjecture.

For the reverse implication, we also use Proposition 23.3. Consider the capped grope G^c represented by the Kirby diagram corresponding to a given link in $\mathcal{L}_1$. In other words, place dots on all components other than $m(L)$, and recall that $m(L)$ is the attaching circle (and does not correspond to a handle). Then the capped grope G^c is the complement of standard slice discs bounded by the unlink formed by the components of $L \smallsetminus m(L)$. As noted earlier, since at least two rounds of ramified Bing doubling are required in the construction of elements in $\mathcal{L}_1$, the grope G^c has height at least two. Contract to produce a grope $G^c(1.5)$ of height 1.5. The latter has the same attaching region as G^c. By Proposition 23.3, the disc embedding conjecture implies Conjecture G, which in turn implies that the attaching circle of $G^c(1.5)$ bounds a flat embedded disc in $G^c(1.5) \subseteq G^c$. By the correspondence described in the first paragraph, this implies Conjecture $\mathcal{L}_1$. □

23.2. The surgery conjecture

A key implication of the disc embedding theorem is the existence and exactness of the surgery sequence, for good fundamental groups, as shown in Chapter 22. The disc embedding conjecture would similarly imply a generalisation of our work in Chapter 22. We present some definitions before precisely stating the surgery conjecture.

Recall from Section 22.2 that the input for the surgery programme for compact, topological 4- manifolds is a 4-dimensional Poincaré pair $(X, \partial X)$, together with a fixed degree one normal map $h\colon N \to \partial X$, where N is a closed 3-manifold. The map h is required to be a simple $\mathbb{Z}[\pi_1(X)]$-coefficient chain homotopy equivalence. In this chapter, we will often work with $(X, \partial X)$ where $\pi_1(X)$ is free, in which case it will suffice for h to be a degree one normal map which is a $\mathbb{Z}[\pi_1(X)]$-homology equivalence (see Remark 22.2). Recall as well that a degree one normal map over X, relative to h, is by definition a quadruple (M, f, ξ, b), where M is a compact, topological 4-manifold with boundary N, the map $f\colon M \to X$ has degree one and restricts to h on the boundary, and (ξ, b) is stable bundle data covering f and restricting to the given data over h.

DEFINITION 23.6. Let $(X, \partial X)$ be a Poincaré pair with $h\colon N \to \partial X$ a degree one normal map from a closed 3-manifold N, inducing a simple $\mathbb{Z}[\pi_1(X)]$-coefficient chain homotopy equivalence. Given a degree one normal map (M, f, ξ, b) over $(X, \partial X)$ relative to h, suppose that $\pi_1(f)$ is an isomorphism, that the surgery kernel $K_2(f)$ is free, and that L is a simple lagrangian for the intersection form $(K_2(f), \lambda, \mu)$.

We say we can do *surgery to kill* L if there exists a degree one normal bordism (W, F, Ξ, B) over $(X, \partial X)$ from (M, f, ξ, b) to some (M', f', ξ', b'), relative to h on the boundary, such that $L = \ker(K_2(f) \to K_2(F))$ and the maps $\pi_1(F)$ and $\pi_1(f)$ are isomorphisms.

Given a degree one normal map (M, f, ξ, b) over X relative to h, assume we have made $\pi_1(f)$ an isomorphism by initial surgeries on embedded circles in M. The surgery kernel is then in general only stably free. Denote by (M_k, f_k, ξ_k, b_k) the degree one normal map that is the effect of performing k surgeries on (M, f, ξ, b) along trivially embedded circles in M. Observe that $M_k = M \#^k S^2 \times S^2$ and that the new intersection form is $(K_2(f_k), \lambda_k, \mu_k) := (K_2(f), \lambda, \mu) \oplus H(\mathbb{Z}[\pi_1(X)]^k)$. After k surgeries, for some k, the surgery kernel becomes free and then the class of the intersection form in $L_4^s(\mathbb{Z}[\pi_1(X)])$ determines the surgery obstruction. This obstruction vanishes if and only if after possibly some further stabilisations of the intersection form by a hyperbolic form, there is a simple lagrangian. Thus the following is exactly what is required to have vanishing surgery obstruction.

DEFINITION 23.7. Let $(X, \partial X)$ be a Poincaré pair with $h\colon N \to \partial X$ a degree one normal map from a closed 3-manifold N, inducing a simple $\mathbb{Z}[\pi_1(X)]$-coefficient chain homotopy equivalence. Given a degree one normal map (M, f, ξ, b) over $(X, \partial X)$ relative to h with $\pi_1(f)$ an isomorphism, a *stable simple lagrangian* for $(K_2(f), \lambda, \mu)$ is a simple lagrangian L for some stabilisation $(K_2(f_k), \lambda_k, \mu_k)$ as defined above.

We now state the the surgery conjecture.

CONJECTURE 23.8 (Surgery conjecture with group π). *Let $(X, \partial X)$ be a 4-dimensional Poincaré pair with $\pi_1(X) =: \pi$ and $h\colon N \to \partial X$ a degree one normal map, where N is a closed 3-manifold, inducing a simple $\mathbb{Z}[\pi]$-coefficient chain homotopy equivalence. Suppose (M, f, ξ, b) is a degree one normal map over $(X, \partial X)$ relative to h, with vanishing surgery obstruction $\sigma(M, f, \xi, b) = 0 \in L_4^s(\mathbb{Z}[\pi])$. Then given any stable simple lagrangian L for $(K_2(f), \lambda, \mu)$, it is possible to do surgery to kill L.*

Chapter 22 shows that the disc embedding conjecture implies the surgery conjecture. The definition in Section 22.1.6 of the leftmost arrow, namely the action of the surgery obstruction group on the structure set, uses Wall realisation, which uses the (category preserving) sphere embedding theorem and thus requires $\pi_1(X)$ to be good. In general, we do not know that this Wall realisation action is defined. However, the surgery conjecture can be used to define this action, as we now show. As a result, the surgery conjecture as stated above implies that the surgery sequence from Chapter 22 is defined and exact.

PROPOSITION 23.9. *Assume the surgery conjecture with group π. Let $(X, \partial X)$ be a 4-dimensional Poincaré pair with $\pi_1(X) \cong \pi$ and $h\colon N \to \partial X$ a degree one* *normal map, where N is a closed* 3*-manifold, inducing a simple* $\mathbb{Z}[\pi_1(X)]$*-coefficient chain homotopy equivalence.*

Then the Wall realisation action of $L_5^s(\mathbb{Z}[\pi_1(X)])$ *on* $\mathcal{S}^s(X,h)$ *is defined, and the resulting surgery sequence*

$$L_5^s(\mathbb{Z}[\pi_1(X)]) \to \mathcal{S}^s(X,h) \to \mathcal{N}(X,h) \xrightarrow{\sigma} L_4^s(\mathbb{Z}[\pi_1(X)]).$$

is exact.

PROOF. The (category preserving) sphere embedding theorem was used in the process of Wall realisation in Section 22.1.5 to construct a normal bordism. This precise bordism is provided by the surgery conjecture instead. Effectively, this means that the surgery conjecture implies Wall realisation and so the proofs that the resulting action of $L_5^s(\mathbb{Z}[\pi_1(X)])$ on $\mathcal{S}^s(X,h)$ is well defined and the exactness of the surgery sequence at the structure set given in Section 22.1.6 still apply. The surgery conjecture also makes the application of the sphere embedding theorem in Section 22.1.4 unnecessary, by once again providing an alternate route to the desired normal bordism, and exactness at the normal maps follows from the same proof as explained there. □

REMARK 23.10. While following through the argument in Section 22.1.5 to recover Wall realisation using the surgery conjecture with group π, as pointed out by Diarmuid Crowley and Jim Davis, it is essential that we have the ability to kill a *specific* lagrangian. By contrast, when following through the argument in Section 22.1.4 to obtain exactness at the normal invariants, the *prima facie* weaker assumption that *some* lagrangian can be killed is all that is required. However, if π is a free group, as illustrated in Figure 23.3, the seemingly weaker statement is equivalent to the surgery conjecture for all groups.

The surgery conjecture is of substantial interest independently from the disc embedding conjecture. In this section, we collect some known equivalences and implications of the surgery conjecture. These are summarised in Figure 23.3.

23.2.1. Good boundary links. For the rest of this chapter we fix some conventions. For a link $L \subseteq S^3$, the result of 0-framed Dehn surgery on S^3 along the components of L is denoted by M_L. We also fix once and for all a preferred isomorphism $\pi_1(\#_{i=1}^n S^1 \times S^2) \cong F_n$, where F_n is the free group on n generators.

Define a *boundary link* (L,ϕ) to be an n-component link $L \subseteq S^3$ together with a surjection $\phi\colon \pi_1(S^3 \setminus L) \twoheadrightarrow F_n$, sending a set of meridians of L to a free generating set. The name derives from the fact, due to Smythe [**Smy66**] (see also [**CS80**, Proposition 1.1]), that a link L admits such a homomorphism ϕ if and only if the components of L bound pairwise disjoint Seifert surfaces in S^3. This equivalence is readily seen: given a collection $V = \sqcup_{i=1}^n V_i$ of mutually disjoint Seifert surfaces for the components of L, a map to $\bigvee_n S^1$ can be constructed by mapping each tubular neighbourhood $V_i \times (-1,1)$ to a copy of the interval $(-1,1)$, and everything else to the wedge point. On the other hand any map ϕ corresponds to a map $S^3 \setminus L \to \bigvee_n S^1$, and mutually disjoint Seifert surfaces can be found by pulling back a collection of regular values, one on each copy of S^1. Observe that given a boundary link (L,ϕ), the function ϕ extends to a function $\phi\colon \pi_1(M_L) \twoheadrightarrow F_n$. This follows since $\pi_1(M_L)$ is the quotient of $\pi_1(S^3 \setminus L)$ by the longitudes of L, and the longitudes are mapped to zero under ϕ.

We will now impose an additional condition to define a certain class of boundary links, called *good boundary links*. It is straightforward from the definition below that a 1-component good boundary link is a knot with Alexander polynomial one.

DEFINITION 23.11. A boundary link (L,ϕ) is said to be a *good boundary link* if the kernel of ϕ is a perfect group.

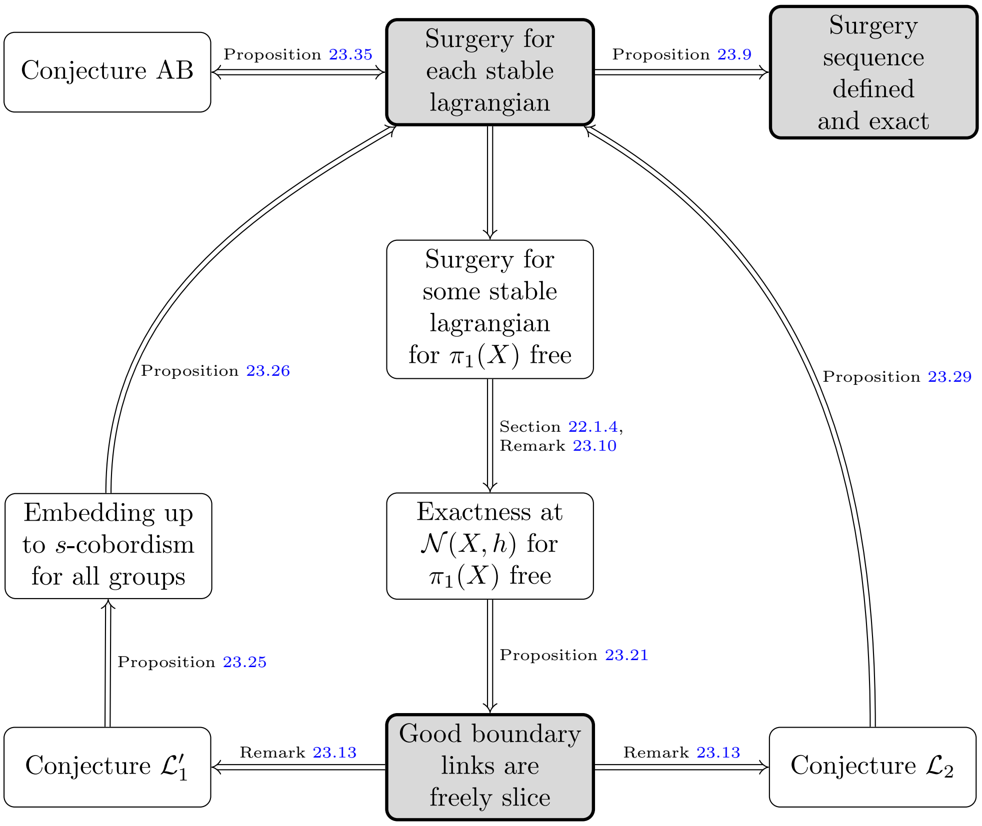


Figure 23.3. The surgery conjecture in relation to other conjectures. Above, $(X, \partial X)$ is a 4-dimensional Poincaré pair.

Note that if (L, ϕ) is a good boundary link then any other choice (L, ϕ') is a good boundary link, since the maps ϕ and ϕ' differ by an automorphism of F_n.

Next we next give a characterisation of good boundary links. Let V denote a collection of pairwise disjoint Seifert surfaces for the components of a boundary link L. Recall that the *Seifert form* $\theta\colon H_1(V;\mathbb{Z}) \times H_1(V;\mathbb{Z}) \to \mathbb{Z}$ is defined via linking numbers of curves on V and their push-offs in $S^3 \smallsetminus V$. The Seifert form is represented by a *Seifert matrix*. A collection $\{a_i, b_i\}_{i=1}^n$ of simple closed curves on V is said to be a *good basis* for V if:

(i) it represents a basis for $H_1(V;\mathbb{Z})$;
(ii) the geometric intersections are $a_i \cdot a_j = 0$, $b_i \cdot b_j = 0$, and $a_i \cdot b_j = \delta_{ij}$ for all $i, j = 1, \dots, n$; and
(iii) the Seifert matrix of V with respect to the basis is of the form

$$\left[\begin{array}{cc:cc:c:cc}
0 & \varepsilon_1 & 0 & * & \cdots & 0 & * \\
1-\varepsilon_1 & 0 & 0 & * & & 0 & * \\
\hdashline
0 & 0 & 0 & \varepsilon_2 & \cdots & 0 & * \\
* & * & 1-\varepsilon_2 & 0 & & 0 & * \\
\hdashline
\vdots & & \vdots & & \ddots & \vdots & \\
\hdashline
0 & 0 & 0 & 0 & \cdots & 0 & \varepsilon_n \\
* & * & * & * & & 1-\varepsilon_n & 0
\end{array}\right]$$

where $\varepsilon_i = 0$ or 1 for each i, and each $*$ designates an arbitrary integer entry.

PROPOSITION 23.12 ([**CKP20**, Corollary 2.2], see also [**FQ90**, Lemma 12.3B]). *A boundary link is good if and only if the components bound a collection of pairwise disjoint Seifert surfaces in S^3 admitting a good basis.*

REMARK 23.13. Each element of $\mathcal{L}_1$ is a good boundary link. To see this we use Proposition 23.12. One may easily construct Seifert surfaces for the links in $\mathcal{L}_1$ so that the form for each component is S-equivalent to that of the unknot. We invite the reader to carry this out for the link shown in Figure 23.2. By the same construction, it is easy to see that any ramified Whitehead double of a link with trivial pairwise linking numbers is a good boundary link.

We saw in Theorem 1.14 that Alexander polynomial one knots are topologically slice, using a surgery-theoretic argument. (See Section 21.6.3 for a different argument using the disc embedding theorem directly.) The next proposition says that every good boundary link can be used to construct a surgery problem where the target is the exterior of a collection of standard slice discs for the unlink, and the surgery obstruction vanishes. From one perspective this is the sole purpose of the definition of a good boundary link.

PROPOSITION 23.14. *Let (L,ϕ) be an n-component good boundary link. Then M_L is the boundary of a compact, topological 4-manifold W such that there exists a degree one normal map*

$$(f,h)\colon (W,M_L) \to (\natural_{i=1}^n S^1\times D^3, \#_{i=1}^n S^1\times S^2),$$

such that h is a $\mathbb{Z}[F_n]$-homology equivalence and ϕ is the composition

$$\pi_1(S^3\smallsetminus L) \to \pi_1(M_L) \xrightarrow{h_*} \pi_1(\#_{i=1}^n S^1\times S^2) \cong F_n,$$

where the first map is induced by inclusion and the last is the preferred isomorphism. Moreover, there exists such a pair (f,h) with vanishing surgery obstruction $\sigma(g)\in L_4(\mathbb{Z}[F_n])$.

In Proposition 23.14, we only claim that h is a $\mathbb{Z}[F_n]$-homology equivalence rather than inducing a simple $\mathbb{Z}[F_n]$-coefficient chain homotopy equivalence. This is sufficient to define a surgery obstruction by Remark 22.2.

We will break the proof of Proposition 23.14 into a series of lemmas, some of which will be reused later.

LEMMA 23.15. *Let X be a Poincaré complex and let N be a compact, smooth manifold. Let $h\colon N\to X$ be a degree one map which is in addition a $\mathbb{Z}$-homology equivalence. Then there exists a stable vector bundle $\xi\to X$ and a vector bundle isomorphism $b\colon h^*\xi\cong\nu_N$. In other words, h may be promoted to a normal map.*

PROOF. We compare the (smooth) vector bundles over N to those over X. Recall that for a pointed space S, the reduced real topological K-theory $\widetilde{KO}(S)$ is the group of stable equivalence classes of real vector bundles with base S. The Atiyah-Hirzebruch spectral sequence applied to the reduced real topological K-theory of cone(h) has $E_2^{p,q}\cong H^p(\mathrm{cone}(h);\widetilde{KO}^q(\mathrm{pt}))$ and converges to $\widetilde{KO}^{p+q}(\mathrm{cone}(h))$. Since h is an integral homology equivalence, and $\widetilde{KO}^q(\mathrm{pt})$ is one of $\mathbb{Z}$, $\mathbb{Z}/2$, or 0 (depending on q), the E_2 page of the sequence vanishes and thus the reduced KO groups of cone(h) are trivial. By the long exact sequence of h in reduced KO theory, h induces an isomorphism $\widetilde{KO}(X)\cong\widetilde{KO}(N)$. It follows that we can find a (stable) vector bundle $\xi\to M$ and an isomorphism $b\colon h^*\xi\cong\nu_N$, as required. $\square$

LEMMA 23.16. *Let (L, ϕ) be an n-component good boundary link. Then there exists a degree one normal map $h\colon M_L \to \#_{i=1}^n S^1 \times S^2$ that is a $\mathbb{Z}[F_n]$-homology equivalence, such that ϕ coincides with the composition*

$$\pi_1(S^3 \smallsetminus L) \to \pi_1(M_L) \xrightarrow{h_*} \pi_1(\#_{i=1}^n S^1 \times S^2) \cong F_n,$$

where the first map is induced by inclusion and the last map is the preferred isomorphism.

PROOF. We appeal to [**CS80**, Proposition 2.2], where obstruction theory is used to show that the exterior of an open tubular neighbourhood of an n-component boundary link admits a map to the exterior of an open tubular neighbourhood of the n-component unlink, preserving longitudes and meridians on the boundary tori, inducing ϕ on fundamental groups, and inducing isomorphisms on integral homology groups. This map can be extended by the identity over the surgery solid tori in the 0-surgery for both L and the unlink, to obtain $h\colon M_L \to \#_{i=1}^n S^1 \times S^2$. The resulting map h is degree one and induces isomorphisms on integral homology. By Lemma 23.15, h moreover extends to a degree one normal map. Note that since h induces ϕ on fundamental groups and the latter has perfect kernel, h is moreover a $\mathbb{Z}[F_n]$-homology equivalence. To see this, we need to compute the homology of M_L with $\mathbb{Z}[F_n]$ coefficients. We have the short exact sequence $0 \to \ker\phi \to \pi_1(M_L) \to F_n \to 0$, and $\ker\phi = \pi_1(\widetilde{M}_L)$, where $\widetilde{M}_L$ is the associated covering space. By assumption $\ker\phi$ is perfect. Therefore $H_1(M_L;\mathbb{Z}[F_n]) \cong \pi_1(\widetilde{M}_L)/\pi_1(\widetilde{M}_L)^{(1)} \cong \{0\}$. By Poincaré duality and the universal coefficient spectral sequence (see, for example, [**Lev77**, Theorem 2.3]), $H_2(M_L;\mathbb{Z}[F_n]) \cong \operatorname{Ext}^1_{\mathbb{Z}[F_n]}(\mathbb{Z}, \mathbb{Z}[F_n])$, which is independent of L. Note that $M_U \cong \#_{i=1}^n S^1 \times S^2$. It then follows from naturality that the induced map $H_2(M_L;\mathbb{Z}[F_n]) \to H_2(\#_{i=1}^n S^1 \times S^2;\mathbb{Z}[F_n])$ is an isomorphism. This completes the proof that h is a $\mathbb{Z}[F_n]$-homology equivalence. □

LEMMA 23.17. *Let (L, ϕ) be an n-component good boundary link. Then for any choice of spin structure on M_L, the class of (M_L, ϕ) vanishes in $\Omega_3^{spin}(BF_n)$.*

PROOF. Fix a choice of spin structure on M_L and write $[(M_L, \phi)] \in \Omega_3^{spin}(BF_n)$ for the corresponding spin bordism class, which we will now show is trivial. Consider the maps $p_i \circ \phi\colon \pi_1(M_L) \to F_n \to \mathbb{Z}$ where p_i is the projection sending the ith generator of F_n (corresponding to a meridian of the ith component L_i of L) to 1 and all other generators to 0. The direct sum of these projection maps induces the isomorphism

$$\Omega_3^{spin}(BF_n) \cong \bigoplus_{i=1}^n \Omega_3^{spin}(B\mathbb{Z}) \cong \bigoplus^n \mathbb{Z}_2.$$

Fix $1 \leq j \leq n$. Observe that we can obtain M_L by first performing 0-framed Dehn surgery on L_j and then the remaining $n-1$ 0-framed Dehn surgeries on the remaining link components. The trace of these $n-1$ surgeries is a cobordism W from M_{L_j} to M_L. This cobordism is spin since L is a boundary link and so has trivial pairwise linking numbers. Moreover the map $p_j \circ \phi$ and the abelianisation $a_j\colon \pi_1(M_{L_j}) \to \mathbb{Z}$ extend over the bordism, so in fact $[(M_L, p_j \circ \phi)] = [(M_{L_j}, a_j)] \in$

$\Omega_3^{spin}(B\mathbb{Z})$, via the commutative diagram

$$\begin{tikzcd}
\pi_1(M_L) \arrow[r, two heads, "\phi"] \arrow[d] & F_n \arrow[d, two heads, "p_j"] \\
\pi_1(W) \arrow[r, two heads] & \mathbb{Z} \\
\pi_1(M_{L_j}) \arrow[u] \arrow[ur, two heads, "a_j"'] &
\end{tikzcd}$$

where the vertical maps on the left are induced by inclusion. It is well known for knots in S^3 that under the isomorphism $\Omega_3^{spin}(B\mathbb{Z}) \cong \Omega_2^{spin} \cong \mathbb{Z}_2$, the spin bordism class of the 0-surgery of the knot, together with its abelianisation map, is equal to the Arf invariant of the knot. In other words, under the isomorphisms $\Omega_3^{spin}(BF_n) \cong \bigoplus_{i=1}^n \Omega_3^{spin}(B\mathbb{Z}) \cong (\mathbb{Z}_2)^n$,

$$[(M_L, \phi)] \mapsto [(M_{L_1}, a_1), \cdots, (M_{L_k}, a_k)] \mapsto (\mathrm{Arf}(L_1), \cdots, \mathrm{Arf}(L_k)).$$

However, each component of a good boundary link has trivial Arf invariant. The latter follows from the commutative diagram

$$\begin{tikzcd}
\pi_1(S^3 \smallsetminus L) \arrow[r, two heads, "\phi"] \arrow[d, two heads] & F_n \arrow[d, two heads, "p_j"] \\
\pi_1(S^3 \smallsetminus L_j) \arrow[r, two heads, "a_j"] & \mathbb{Z}
\end{tikzcd}$$

where the left vertical arrow is induced by setting the meridians for all components other than L_j to zero, and a_j is still the abelianisation map by a slight abuse of notation. The kernel of a_j is a quotient of the kernel of ϕ, so is perfect by hypothesis. Thus L_j has trivial Alexander polynomial, which implies that it has trivial Arf invariant since the Arf invariant vanishes if $\Delta_{L_j}(-1) = \pm 1 \mod 8$ by [**Lev66**]. Consequently, $[(M_L, \phi)]$ is trivial in $\Omega_3^{spin}(BF_n)$ as claimed. So (M_L, ϕ) bounds a smooth null bordism over BF_n. □

REMARK 23.18. For a good boundary link L, there is also a direct method to construct a smooth, spin 4-manifold over F_n with boundary M_L, using a boundary link Seifert surface V for L. Attach 0-framed 2-handles to D^4 along the components of L and denote the result by X_L. Note this is spin since L has trivial pairwise linking numbers as a boundary link. Let W denote the result of surgery on X_L along the components of the (framed) capped off Seifert surface $\widehat{V}$, after pushing the interior of V into D^4. That is, $W := (X_L \smallsetminus \nu\widehat{V}) \cup R \times S^1$ where R is a union of 3-dimensional handlebodies such that ∂R is homeomorphic to $\widehat{V}$. Then W is spin since the Arf invariants of L vanish. In more detail, the Arf invariant of a component L_i determines whether the induced spin structure on the component $\widehat{V}_i$ extends over some 3-dimensional handlebody R_i with $\partial R_i = \widehat{V}_i$. Since all the Arf invariants vanish, the union of handlebodies $R := \bigsqcup R_i$ exists, with spin structures extending those on the $\widehat{V}_i$. Since $\pi_1(W) \cong F_n$ is normally generated by the meridians of L, the map $\phi\colon \pi_1(M_L) \to F_n$ extends to $\pi_1(W)$.

LEMMA 23.19. *Let (L, ϕ) be an n-component boundary link and suppose there exists a map h as described in Lemma 23.16 and a smooth null bordism (W, Φ) of*

(M_L, ϕ) over F_n. Then there is a degree one normal map

$$(f,h)\colon (W, M_L) \to (\natural_{i=1}^n S^1 \times D^3, \#_{i=1}^n S^1 \times S^2),$$

inducing Φ on fundamental groups if and only if W is spin.

Proof. The "only if" direction is relatively easy. The stable normal bundle of $\natural_{i=1}^n S^1 \times D^3$ is framable, so a choice of framing can be pulled back using the normal data covering g to determine a framing of the stable normal bundle of W. This determines a framing of the stable tangent bundle of W, which in turn determines a spin structure on W. Indeed, since the natural map $BO(4) \to BO$ is 4-connected, the trivialisation of the stable tangent bundle on the 2-skeleton determines a trivialisation of the unstable tangent bundle, whose restriction to the 1-skeleton yields a spin structure.

For the converse, assume that W is spin. In order to construct f, observe that $i \circ h\colon M_L \to \#_{i=1}^n S^1 \times S^2 \hookrightarrow \natural_{i=1}^n S^1 \times D^3$ equals ϕ on the level of fundamental groups, where i is the inclusion of the boundary. Since $\natural_{i=1}^n S^1 \times D^3$ is a classifying space BF_n, we can choose a map $f\colon W \to \natural_{i=1}^n S^1 \times D^3$ representing $\Phi\colon \pi_1(W) \to F_n$. Moreover, as ϕ and Φ are compatible under the inclusion $M_L \subseteq W$, the map f can be chosen to be compatible with the composition $i \circ h$. The resulting map f has degree one because h has degree one. We have the following diagram of spaces

$$\begin{tikzcd} M_L \arrow[r, "h"] \arrow[d] & \#_{i=1}^n S^1 \times S^2 \arrow[d, "i"] \\ W \arrow[r, "f"] & \natural_{i=1}^n S^1 \times D^3. \end{tikzcd}$$

We must show that the normal data covering h can be extended to f. For this, note that the choice of spin structure on W also determines a choice of trivialisation of ν_W (this is again the fact that for an oriented 4-manifold with boundary a choice of spin structure is equivalent to a choice of framing). The choice of spin structure restricts to a spin structure on M_L, and thus a choice of trivialisation of ν_{M_L}. Since h is a normal map, this induces a choice of trivialisation on the stable normal bundle of $\#_{i=1}^n S^1 \times S^2$. Extend this to a trivialisation for the stable normal bundle of $\natural_{i=1}^n S^1 \times D^3$, so we now have a trivial stable vector bundle $\xi \to \natural_{i=1}^n S^1 \times D^3$. Now take $b\colon \nu_W \to \xi$, the map determined by the original trivialisation of ν_W. This choice of normal data (ξ, b) covers f and has the required restriction on the boundary. □

Remark 23.20. Generally, given a 4-dimensional Poincaré pair $(X, \partial X)$ and a degree one normal map $h\colon N \to \partial X$ from a closed 3-manifold N, the obstruction to the existence of a degree one normal map $(f,h)\colon (W,N) \to (X, \partial X)$, relative to h, lies in the group $H^3(X, \partial X; \mathbb{Z}/2\mathbb{Z}) \cong H_1(X; \mathbb{Z}/2\mathbb{Z})$. We have shown that when $(X, \partial X) = (\natural_{i=1}^n S^1 \times D^3, \#_{i=1}^n S^1 \times S^2)$ and $N = M_L$ for a good boundary link L this obstruction group is in fact $\Omega_3^{spin}(BF_n)$, with the obstruction being given by the class of (M_L, ϕ).

Proof of Proposition 23.14. Use Lemma 23.16 to obtain a map h as described there. Now use Lemma 23.17 to obtain a smooth, spin null bordism (W, Φ) of (M_L, ϕ) over F_n. Apply Lemma 23.19 to h and (W, Φ) to obtain a degree one normal map

$$(f,h)\colon (W, M_L) \to (\natural_{i=1}^n S^1 \times D^3, \#_{i=1}^n S^1 \times S^2)$$

inducing Φ on fundamental groups. This surgery problem may have nontrivial surgery obstruction $\sigma(g) \in L_4(\mathbb{Z}[F_n])$. To modify this, recall that it follows from Cappell's splitting theorem [**Cap71**] that $L_4(\mathbb{Z}[F_n]) \cong L_4(\mathbb{Z}) \cong 8\mathbb{Z}$, with the first isomorphism induced by the augmentation $\mathbb{Z}[F_n] \to \mathbb{Z}$ and the second given by the signature. Thus, we may use connected sums with Freedman's E_8-manifold, possibly with reversed orientation, to modify W so that f has trivial surgery obstruction. This completes the proof. $\square$

The following proposition shows that the analogue of Theorem 1.14, that Alexander polynomial one knots are topologically slice, holds for good boundary links, assuming the exactness of the surgery sequence at the normal maps for free fundamental group. The proof is virtually the same as for Theorem 1.14; the only difference is that $\mathbb{Z}$ is known to be good, while non-abelian free groups are not (see Chapter 19).

PROPOSITION 23.21. *Let $(X, \partial X)$ be a 4-dimensional Poincaré pair with $h\colon N \to \partial X$ a degree one normal map, where N is a closed 3-manifold, inducing a simple $\mathbb{Z}[\pi_1(X)]$-coefficient chain homotopy equivalence. Assume that $\pi_1(X)$ is free. Suppose the surgery sequence is exact at $\mathcal{N}(X, h)$. Then every good boundary link in S^3 is freely topologically slice in D^4.*

Above, we say that a link is *freely topologically slice* in D^4 if it has a collection of locally flat pairwise disjoint slice discs in D^4 whose complement has free fundamental group, freely generated by the meridians of the link. Such a collection of slice discs is said to be *free*.

PROOF. Let L be an n-component good boundary link. By Proposition 23.14, there exists a degree one normal map

$$(f, h)\colon (W, M_L) \to (\natural_{i=1}^n S^1 \times D^3, \#_{i=1}^n S^1 \times S^2)$$

which is a $\mathbb{Z}[F_n]$-homology equivalence on the boundary and has vanishing surgery obstruction $\sigma(f) \in L_4(\mathbb{Z}[F_n])$. By the hypothesis, W is normally bordant, relative to the boundary, to a 4-manifold V which is homotopy equivalent to $\natural_{i=1}^n S^1 \times D^3$. Now attach n copies of $D^2 \times D^2$ to V along the meridians of L in $M_L = \partial V$. The resulting compact, topological 4-manifold X is contractible and has boundary S^3. The 4-dimensional topological Poincaré conjecture (Section 21.6.2) then implies that X is homeomorphic to D^4. The cocores of the attached 2-handles are slice discs for the components of L. Note that V is the exterior of these slice discs and has free fundamental group freely generated by the meridians of L. In other words, L is freely topologically slice in D^4. $\square$

23.2.2. Free slice discs and the link family $\mathcal{L}_1$. We have shown that the surgery conjecture implies that all good boundary links are freely topologically slice. This only requires a consequence of the surgery conjecture, namely exactness of the sequence at the normal maps for free fundamental groups. In fact, the sliceness of good boundary links implies the surgery conjecture. We describe two paths for this implication, as shown in Figure 23.3. The first involves the family of links $\mathcal{L}_1$ from the previous section.

CONJECTURE 23.22 (Conjecture $\mathcal{L}_1'$). *Each link L in the family $\mathcal{L}_1$ is topologically slice in D^4 where the sub-collection of slice discs for $L \setminus m(L)$ is free.*

Since the elements of $\mathcal{L}_1$ are good boundary links by Remark 23.13, we see that the surgery conjecture implies Conjecture $\mathcal{L}_1'$ by Proposition 23.21.

Recall that removing $m(L)$ from a link $L \in \mathcal{L}_1$ results in the unlink. Thus the next property of free slice discs for the unlink will be very useful.

LEMMA 23.23. *The exterior of any collection of free slice discs for the n-component unlink is s-cobordant relative to the boundary to $\natural_{i=1}^n S^1 \times D^3$.*

Proof. Let X be the exterior of a collection of free slice discs for the n-component unlink and note that $\partial X = \#_{i=1}^n S^1 \times S^2$. Elementary obstruction theory implies that the identity map on $\#_{i=1}^n S^1 \times S^2$ extends to a map $f\colon X \to \natural_{i=1}^n S^1 \times D^3$ that is an isomorphism on π_1. Note that f realises the $\mathbb{Z}$-homology equivalence guaranteed by Alexander duality for the slice discs in D^4. A calculation [**FQ90**, Theorem 11.6C(1)] shows that f is moreover a $\mathbb{Z}[F_n]$-homology equivalence, so by Whitehead's theorem it is a homotopy equivalence. Since f is degree one on the boundary, it is degree one as a map of pairs

$$(f, \mathrm{Id})\colon (X, \partial X) \to (Y, \partial Y) := (\natural_{i=1}^n S^1 \times D^3, \#_{i=1}^n S^1 \times S^2).$$

We want to upgrade f to a normal map, but we cannot apply Lemma 23.19 directly since X is not a smooth null bordism. Instead, we argue as follows. Since X is a subset of D^4, the topological tangent bundle is trivial. Let ξ be a trivial bundle over Y. Then the sum $TX \oplus f^*\xi$ is a trivial bundle. Choose a trivialisation of it, which determines a stable trivialisation. This gives a degree one normal map in Wall's sense [**Wal99**, pp. 9–10], since a stable trivialisation of $TX \oplus f^*\xi$ is equivalent to an isomorphism $\nu_X \cong f^*\xi$, that is a normal map. So f can be upgraded to a normal map as desired.

The set of degree one normal maps, relative to the boundary, to $(Y, \partial Y)$ is given by the set of homotopy classes

$$[(Y/\partial Y), G/TOP] = H^2(Y, \partial Y; \mathbb{Z}/2\mathbb{Z}) \oplus H^4(Y, \partial Y; \mathbb{Z}) \cong 0 \oplus \mathbb{Z}$$

(compare with Section 22.3.2). The integer in the latter summand is given by taking the image of $\sigma(f)$ under the map induced by augmentation $L_4(\mathbb{Z}[F_n]) \to L_4(\mathbb{Z}) \cong 8\mathbb{Z}$, and then dividing by 8. Since f is a homotopy equivalence, $\sigma(f)$ is already 0, and so this integer is 0. Thus there exists a degree one normal bordism, relative to the boundary, from $f\colon X \to Y$ to the identity map $\mathrm{Id}\colon Y \to Y$.

Call the degree one normal bordism just obtained $F\colon U \to Y \times [0,1]$. We will modify the 5-manifold U to be an s-cobordism over $Y \times [0,1]$. First, note that a corollary of [**BHS64**, Theorem 2] is that the Whitehead group of a free group is trivial, so in fact there is no distinction between homotopy equivalences and simple homotopy equivalences for our purposes. The degree one normal map F, along with normal data, determines a 5-dimensional surgery problem, with surgery obstruction $\sigma(F) \in L_5(\mathbb{Z}[F_n])$. To identify the surgery obstruction we use Cappell's splitting theorem [**Cap71**] again, to obtain an isomorphism $L_5(\mathbb{Z}[F_n]) \cong \bigoplus_{k=1}^n L_4(\mathbb{Z})$. In order to modify the entry of $\sigma(F)$ in an individual $L_4(\mathbb{Z})$ summand, say the kth summand, take the surgery problem $p\colon S^1 \times E_8 \to S^1 \times S^4$, that is, the canonical degree one normal map obtained as a collapse map where E_8 denotes Freedman's E_8-manifold, and perform a fibre sum

$$F \#_{S^1} p\colon U \#_{S^1} (S^1 \times E_8) \to (Y \times [0,1]) \#_{S^1} \left(S^1 \times S^4\right) \cong Y \times [0,1].$$

We explain the fibre sum operation in more detail. Assume, by attaching a collar if necessary, that there is a neighbourhood of $Y \subseteq \partial U$ in U on which F restricts to the identity. Push off the core $S^1 \times \{*\}$ of the summand of $Y = \natural_{i=1}^n S^1 \times D^3$ corresponding to the kth generator of F_n into this collar. Then to perform the fibre sum on $F\colon U \to Y \times [0,1]$, we identify a trivially framed tubular neighbourhood of this push-off with a trivially framed tubular neighbourhood of $S^1 \times \{*\}$ in $S^1 \times E_8$ in the domain U, and with a trivially framed tubular neighbourhood of $S^1 \times \{*\}$ in $S^1 \times S^4$ in the codomain $Y \times [0,1]$. Finally, observe that fibre sum with $S^1 \times S^4$ as defined preserves homeomorphism type.

By repeatedly performing this operation (or the version with the reversed orientation on the E_8-manifold), kill the surgery obstruction $\sigma(F)$. Since there is no longer a surgery obstruction, we may perform surgeries on the interior of the new U to make F a homotopy equivalence. The resulting 5-manifold will be an s-cobordism (relative to the boundary) from X to $Y = \natural_{i=1}^n S^1 \times D^3$ as required. $\square$

We noted earlier that the surgery conjecture with π a free group would imply Conjecture $\mathcal{L}_1'$ via Proposition 23.21 and Remark 23.13. In fact the converse is also true. The proof of this goes via an intermediate conjecture, also of interest.

Conjecture 23.24 (Disc embedding up to s-cobordism with group π). *Let M be a connected, topological 4-manifold with fundamental group π and nonempty boundary. 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 there is an immersed collection of framed dual spheres

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

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

Then there exists a connected, topological 4-manifold M' that is s-cobordant to M relative to the boundary via a compactly supported s-cobordism W, and there exist mutually disjoint, flat 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 an immersed collection of framed geometrically transverse 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 in $\partial M = \partial M'$ and the sphere $\overline{g}_i$ is homotopic to g_i in W.

Proposition 23.25. *Conjecture $\mathcal{L}_1'$ implies disc embedding up to s-cobordism for all groups.*

Proof. Assume Conjecture $\mathcal{L}_1'$. Let $L \in \mathcal{L}_1$ be an $(n+1)$-component link. Since $L \smallsetminus m(L)$ is the unlink, we may apply Lemma 23.23 to see that the complement in D^4 of the free slice discs for $L \smallsetminus m(L)$ is s-cobordant, relative to the boundary, to $\natural_{i=1}^n S^1 \times D^3$. Via the interpretation of elements of $\mathcal{L}_1$ as Kirby diagrams for capped gropes of height two described in the proof of Proposition 23.5, this means that each disc-like capped grope of height two and with mutually disjoint caps is s-cobordant relative to its boundary to a manifold, namely the complement in D^4 of the free slice discs for $L \smallsetminus m(L)$, for the corresponding link L, where the attaching circle bounds a flat embedded disc. Following the second half of the proof of Proposition 23.3, this statement is sufficient to complete the proof. Specifically, the hypothesised s-cobordisms may be appropriately placed within $M \times [0, 1]$ to obtain the desired s-cobordism from M to some topological manifold M', containing the desired embedded discs. $\square$

Proposition 23.26. *For any group π, disc embedding up to s-cobordism with group π implies the surgery conjecture with group π.*

Proof. Fix a group π. First note that disc embedding up to s-cobordism implies sphere embedding up to s-cobordism, by following the proof of the sphere embedding theorem in Chapter 20. We sketch this proof now. Beginning with spheres $\{f_i\}$ with vanishing intersection and self-intersection numbers, equipped with framed, dual spheres $\{g_i\}$, first we assume they are geometrically transverse by the geometric Casson lemma (Lemma 15.3). Next find framed, immersed Whitney discs $\{D_k'\}$ for the intersections within $\{f_i\}$, with interiors disjoint from $\{f_i\}$ and also equipped with transverse spheres $\{g_k'\}$. The spheres $\{g_k'\}$ are produced by contracting Clifford tori at double points of $\{f_i\}$. Working in the complement of

$\{f_i\}$, apply disc embedding up to s-cobordism to find an s-cobordant manifold in which the Whitney circles bound mutually disjoint, flat embedded discs $\{W_k\}$ with geometrically transverse spheres $\{R_k\}$, so that each R_k is homotopic g_k'. By tubing $\{g_i\}$ into $\{R_k\}$ we produce spheres $\{\overline{g}_i\}$ which are geometrically transverse to $\{f_i\}$ and moreover disjoint from $\{W_k\}$. By Lemma 17.11, each $\overline{g}_i$ is homotopic to g_i. In the new ambient manifold, which is s-cobordant to the original one, the spheres $\{f_i\}$ come with mutually disjoint, framed, embedded Whitney discs, and are therefore regularly homotopic to embeddings. The resulting spheres $\{\overline{f}_i\}$ are geometrically transverse to the collection $\{\overline{g}_i\}$. This establishes sphere embedding up to s-cobordism.

Suppose $(X, \partial X)$ is a 4-dimensional Poincaré pair with $\pi_1(X) = \pi$ and we are given a degree one normal map (M, f, ξ, b) over $(X, \partial X)$, relative to some $h\colon \partial M \to \partial X$, with a given stable simple lagrangian P for the intersection form on $K_2(f)$. Realise the required stabilisation by connected sum with k copies of $S^2 \times S^2$. Now the intersection form on the surgery kernel is isomorphic to the hyperbolic form $H(P)$. By sphere embedding up to s-cobordism, we can produce a manifold $(M', \partial M' = \partial M)$, s-cobordant to $(M \#^k S^2 \times S^2, \partial M)$, in which the relevant homotopy classes for surgery to kill P are regularly homotopic to framed embedded spheres. Attaching 5-dimensional handles along these framed spheres yields a normal bordism. Thus, the proposition will be proved by Lemma 23.27, below. Note that the lemma is part of showing that the map $\mathcal{S}^s(X, h) \to \mathcal{N}(X, h)$ in the (relative) surgery sequence is well defined. □

Lemma 23.27. *Let $(X, \partial X)$ be a 4-dimensional Poincaré pair with a degree one normal map $h\colon N \to \partial X$ inducing a simple $\mathbb{Z}[\pi_1(X)]$-coefficient chain homotopy equivalence, where N is a closed 3-manifold. Then given a degree one normal map (M, f, ξ, b) over $(X, \partial X)$ relative to h, and given a manifold $(M', \partial M')$ that is s-cobordant to $(M, \partial M)$, relative to the boundary $N = \partial M = \partial M'$, there exists a degree one normal map (M', f', ξ', b') degree one normal bordant to (M, f, ξ, b), relative to the boundary.*

Proof. Let $(W, \partial W)$ be an s-cobordism between $(M, \partial M)$ and $(M', \partial M')$, relative to the boundary, that is, $\partial M = \partial M'$, $\partial W = -M \cup_{\partial M} (\partial M \times [0,1]) \cup_{\partial M} M'$, and the inclusions $i\colon M \to W$, $i'\colon M' \to W$ are homotopy equivalences that restrict to the identity maps on ∂M and $\partial M'$ respectively. The maps i, i' have homotopy inverses $j\colon W \to M$, and $j'\colon W \to M'$ respectively, such that $j|_M\colon M \to M$ and $j'|_{M'}\colon M' \to M'$ equal the identity maps. We modify W slightly by adding a collar $M \times [0,1]$ along M, to obtain a new s-cobordism $\overline{W} := (M \times [0,1]) \cup_M W$ between $(M, \partial M)$ and $(M', \partial M')$. This is just for convenience of notation and indeed $\overline{W}$ is homeomorphic to W.

We now construct a degree one normal map $(\overline{W}, F, \Xi, B)$, relative to h, restricting to (M, f, ξ, b) on one end and inducing a degree one normal map (M', f', ξ', b') on the other end. We define the map $F\colon \overline{W} \to X \times [0,1]$ on the collar of $\overline{W}$ by $f \times \mathrm{Id}\colon M \times [0,1] \to X \times [0,1]$. We define F on the remaining part $W \to X \times \{1\} \subseteq X \times [0,1]$ as the composition $f \circ j$ followed by the natural identification $X = X \times \{1\}$. Observe that F restricts to f on the M end of $\overline{W}$ and the induced map $f'\colon M' \to X$ is $f \circ j \circ i'$. On $\partial M' = \partial M$, the map f' restricts to $f|_{\partial M} = h$ as required. This induced map f' is moreover degree one, as $f \circ (j \circ i')\colon M' \to M \to X$ is the composition of a homotopy equivalence relative to the boundary, followed by a degree one map. To see that F is degree one, consider the map of long exact homology sequences of pairs

$$\begin{tikzcd}
0 \arrow[r] \arrow[d] & 0 \arrow[r] \arrow[d] & H_5(\overline{W}, \partial\overline{W}) \arrow[r] \arrow[d, "F"] & H_4(\partial\overline{W}) \arrow[r] \arrow[d, "\cong", "-f\cup(f\times\mathrm{Id})\cup f'"'] & H_4(\overline{W}) \arrow[d, "\cong", "F"'] \\
0 \arrow[r] & 0 \arrow[r] & H_5(X\times[0,1], \partial(X\times[0,1])) \arrow[r] & H_4(\partial(X\times[0,1])) \arrow[r] & H_4(X\times[0,1]).
\end{tikzcd}$$

By the five lemma, the middle map is an isomorphism. A more careful diagram chase, using that f is degree one, shows that F is also degree one.

Finally we need to show that F is a normal map. Recall that since $\overline{W}$ is a manifold, it is equipped with its stable normal bundle $\nu_{\overline{W}}$, which is classified by a map $\overline{W} \xrightarrow{\nu_{\overline{W}}} BTOP$. Since the stable normal bundle on $\overline{W}$ restricts to the stable normal bundle on its boundary, the classifying map for ν_M factors through $\overline{W}$. We have the following diagram:

$$\begin{tikzcd}
M \arrow[dr, shift left, "\overline{i}"] \arrow[drr, bend left, "\nu_M"] \arrow[ddd, "f"] & & \\
& \overline{W} \arrow[ul, shift left, "\overline{j}"] \arrow[r, "\nu_{\overline{W}}"] \arrow[d, "F"] & BTOP \arrow[d, equal] \\
& X\times[0,1] \arrow[r, "\Xi"] \arrow[dl, shift left, "p"] & BTOP \\
X \arrow[ur, shift left, "\iota"] \arrow[urr, bend right, "\xi"'] & &
\end{tikzcd}$$

where $p\colon X\times[0,1] \to X$ is the projection and Ξ is defined as the pullback $p^*\xi$. The map $\iota\colon X \hookrightarrow X\times[0,1]$ is the inclusion identifying X with $X\times\{0\}$. The map $\overline{i}$ is the inclusion $M \hookrightarrow \partial\overline{W}$ and the map $\overline{j}$ is defined on $\overline{W} = W \cup (M\times[0,1])$ as the union $j \cup \mathrm{Id}$ followed by projection onto $M\times\{0\}$. All regions in the diagram above, except for the square on the right, commute by construction; our goal is to show that the square on the right commutes, that is, $\nu_{\overline{W}} \cong F^*\Xi$.

We have that $\nu_{\overline{W}} = \overline{j}^*\nu_M$ and $p^*\xi =: \Xi$, as well as $f^*\xi \cong \nu_M$ by hypothesis. Then we have the sequence of isomorphisms $\nu_{\overline{W}} = \overline{j}^*\nu_M \cong \overline{j}^* f^*\xi = F^*p^*\xi = F^*\Xi$, where the penultimate equality is due to the left trapezoid commuting. This completes the proof. □

23.2.3. Free slice discs and the link family $\mathcal{L}_2$. The surgery conjecture had been investigated prior to Freedman's proof of the disc embedding theorem. In [**CF84**], the surgery conjecture was reformulated in terms of a family of "atomic surgery problems", which may also be rephrased as a universal link slicing problem. These were in terms of Casson towers. We show below how to recast the atomic surgery problems, or rather the related universal link slicing problem, in terms of capped gropes.

Let $\mathcal{L}_2$ denote the class of links obtained by iterated, ramified Bing doubling on either component of the Hopf link followed by a final round of ramified Whitehead doubling on all the components. We require that at least one of the two components of the original Hopf link is subjected to at least one round of ramified Bing doubling. As before for $\mathcal{L}_1$, the ramification parameters need not be constant across components, and all possible clasps are allowed in the final Whitehead doubling step. Iterated, ramified Bing doubles of the Hopf link are called *generalised Borromean rings* (GBRs). The Borromean rings themselves are the simplest GBR, since they are produced by (unramified) Bing doubling of one of the two components of the Hopf link. Observe that each element of $\mathcal{L}_2$ is obtained by performing ramified Whitehead doubling on each component of some GBR. We have the following conjecture and corresponding universal link slicing problem.

CONJECTURE 23.28 (Conjecture $\mathcal{L}_2$). *Each element of the family $\mathcal{L}_2$ is freely topologically slice.*

By Remark 23.13, each element of $\mathcal{L}_2$ is a good boundary link, as a ramified Whitehead double of a link with trivial pairwise linking numbers. So by Proposition 23.21 the surgery conjecture implies Conjecture $\mathcal{L}_2$. Indeed the converse is also true.

PROPOSITION 23.29. *Conjecture $\mathcal{L}_2$ implies the surgery conjecture for all groups.*

The beginning of the proof is similar to our work in Chapter 16, but the details are different, so we go through them. The argument here is closer to the proof outlined in [**FQ90**, Chapter 5, pp. 86-7].

PROOF. Suppose $(X, \partial X)$ is a 4-dimensional Poincaré pair and we are given a degree one normal map (M, f, ξ, b) over X, relative to some fixed $h\colon \partial M \to \partial X$, which is a degree one normal map inducing a simple $\mathbb{Z}[\pi_1(X)]$-coefficient chain homotopy equivalence, as well as a specified stable simple lagrangian P for the intersection form on $K_2(f)$. Since we may realise the required stabilisation by connected sum on M with k copies of $S^2 \times S^2$, by an abuse of notation we assume that the intersection form on the surgery kernel is isomorphic to the hyperbolic form $H(P)$.

Consider framed, immersed spheres $\{f_i\}$ representing the given basis of P and immersed spheres $\{g_i\}$ generating P^*. After adding local kinks, we have that $\lambda(f_i, f_j) = \lambda(g_i, g_j) = 0$, $\mu(f_i) = \mu(g_i) = 0$, and $\lambda(f_i, g_j) = \delta_{ij}$ for all i, j. Recall, as elsewhere in the chapter, we implicitly use the immersion lemma, topological transversality, and the existence of normal bundles for locally flat submanifolds of topological 4-manifolds, in order to work in the topological category.

At this moment, the surgery kernel is represented by $\bigcup\{f_i\} \cup \bigcup\{g_i\}$, namely algebraically transverse collections of immersed spheres. The key step in the proof consists of representing the surgery kernel by geometrically transverse sphere-like capped gropes, with pairwise disjoint caps instead. This is the content of the following lemma. The proof uses techniques from Part II.

LEMMA 23.30. *The surgery kernel may be represented by a collection*

$$\bigcup\{F_i^c\} \cup \bigcup\{G_i^c\},$$

where each F_i^c and G_i^c is a sphere-like capped grope of arbitrary height, the collection $\{F_i^c\}$ represents the given basis of P, the collection $\{G_i^c\}$ represents the given basis of P^*, the collections $\{F_i^c\}$ and $\{G_i^c\}$ are geometrically transverse, and moreover all the caps are mutually disjoint.

PROOF. Apply the geometric Casson lemma (Lemma 15.3) to replace the spheres $\{f_i\}$ and $\{g_i\}$ by geometrically transverse collections $\{f_i'\}$ and $\{g_i'\}$ such that each f_i and g_i is regularly homotopic to f_i' and g_i' respectively. Additionally, each family $\{f_i'\}$ and $\{g_i'\}$ is pairwise disjoint. Since each f_i' is regularly homotopic to f_i, we have that $\lambda(f_i', f_j') = \mu(f_i') = 0$ for all i, j. Thus the intersections and self-intersections within $\{f_i'\}$ can be paired by framed, immersed Whitney discs in M.

Tube along parts of these Whitney circles, as shown in Figure 23.4, to transform $\{f_i'\}$ into a collection $\{\Sigma_i\}$ of pairwise disjoint embedded closed surfaces, with geometrically transverse spheres $\{g_i'\}$.

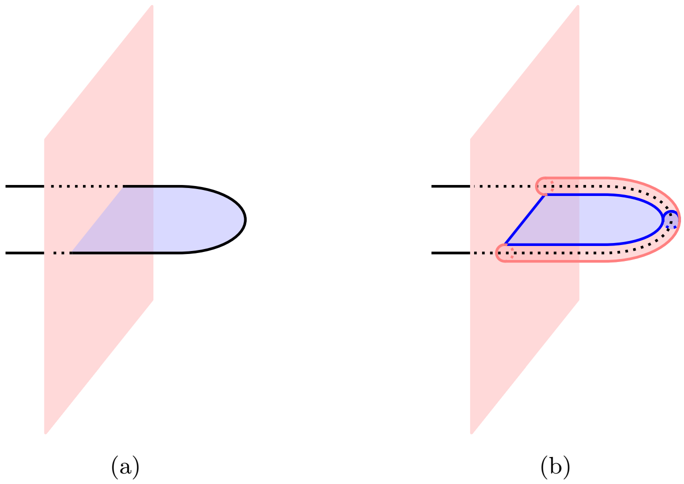


Figure 23.4. Tubing along Whitney circles modifies immersed spheres into embedded, closed, capped surfaces, with caps given by Whitney discs and meridional discs. Black and red above denote elements of the set $\{f_i'\}$, potentially the same element. Note that the Whitney disc (blue) has uncontrolled intersections (not pictured), while the meridional disc (also blue) only intersects $\{f_i'\}$, exactly once.

Cap $\{\Sigma_i\}$ with the Whitney discs from above and meridional discs for $\{f_i'\}$. Each meridional cap intersects $\{\Sigma_i\}$ exactly once while the caps obtained from the Whitney discs may intersect one another, $\{\Sigma_i\}$, or $\{g_i'\}$. Tube the intersection points of the caps with $\{\Sigma_i\}$ into the geometrically transverse spheres $\{g_i'\}$. As a result, the caps no longer intersect $\{\Sigma_i\}$, at the cost of introducing (potentially many) new intersections among caps, and between caps and $\{g_i'\}$. We have replaced $\{f_i\}$ by a union-of-spheres-like capped grope of height one, but the $\{g_i'\}$ are not geometrically transverse, since in particular the $\{g_i'\}$ likely intersect the caps of the capped grope.

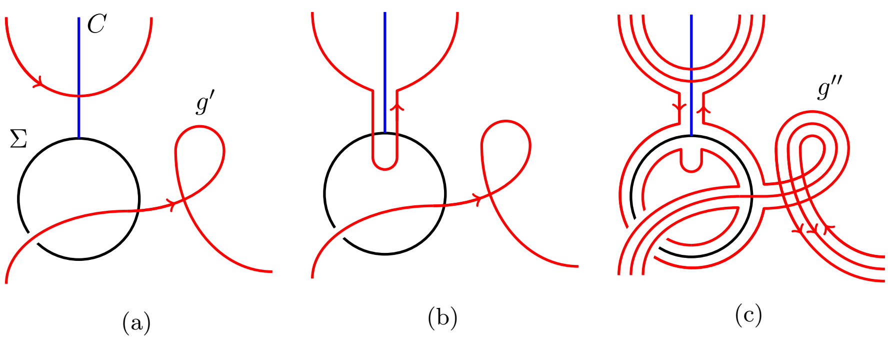


Figure 23.5. Resolving intersections between a sphere g' (red) transverse to a surface Σ (black, shown as a cross section) and its caps (blue). After pushing the intersection between g' and the cap C down to Σ, we tube into the "pre-pushing" g'. The result is called g''.

Next we proceed to make the caps disjoint from $\{g_i'\}$. For each intersection between $\{g_i'\}$ and a cap, push the intersection down to the surfaces $\{\Sigma_i\}$, by pushing g_i'. This results in two points of intersection between $\{g_i'\}$ and $\{\Sigma_i\}$, of cancelling sign. At these two points of intersection, tube the pushed $\{g_i'\}$ into parallel copies of the pre-pushing $\{g_i'\}$; call the resulting set $\{g_i''\}$ (see Figure 23.5). This procedure removes the original intersections between $\{g_i'\}$ and the caps, but creates new self-intersections of $\{g_i''\}$, as well as new intersections of $\{g_i''\}$ with the caps. However, since $\{g_i''\}$ is constructed from $\{g_i'\}$ by tubing an even number of times with opposite orientation, all the new intersections can be paired up with Whitney discs.

As a result, $\mu(g_i'') = \lambda(g_i'', g_j'') = 0$ for all i, j, and moreover, the intersection points between $\{g_i''\}$ and the caps occur in pairs with corresponding Whitney discs, that is, for each cap C, we have $\lambda(C, g_i'') = 0$ for all i. Moreover, the sets $\{g_i''\}$ and $\{\Sigma_i\}$, not including any caps, are geometrically transverse. We are now in good shape to apply the geometric Casson lemma to the caps and $\{g_i''\}$, but we postpone this for a moment while we ensure that the intersections among the caps are also manageable.

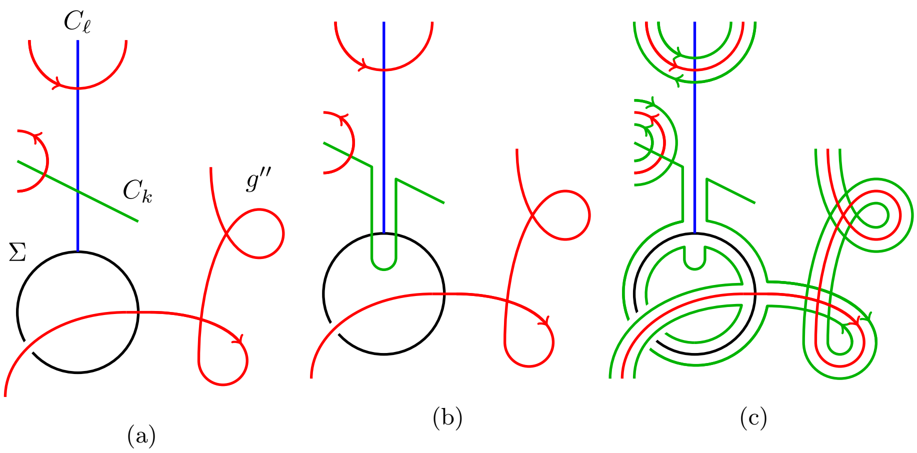


Figure 23.6. Managing cap intersections. Blue and green represent two caps for the surface Σ (black, shown as a cross section). After pushing the green cap down to Σ, we tube into the geometrically transverse sphere g''. Note that all new blue-green intersections occur in algebraically cancelling pairs.

Recall that at the moment a cap may intersect itself as well as any other cap. For each such intersection between two caps (possibly a self-intersection), push the intersection down to the surface Σ and then tube into $\{g_i''\}$ (see Figure 23.6). This will create more intersections between the caps and $\{g_i''\}$, as well as among the caps, but as before, these all occur in algebraically cancelling pairs. We have now achieved that

$$\lambda(C_\ell, C_k) = \mu(C_\ell) = \lambda(C_\ell, g_i'') = 0$$

for every i, k, ℓ, where $\{C_\ell\}$ is a collection of caps for $\{\Sigma_i\}$.

Thus, we have Whitney discs pairing up the intersection points between $\{C_\ell\}$ and $\{g_i''\}$. Consider one such Whitney disc D. This disc may intersect $\{\Sigma_i\}$, any cap, or $\{g_i''\}$. Recall that $\{g_i''\}$ and $\{\Sigma_i\}$ are geometrically transverse. For every point of intersection between $\{\Sigma_i\}$ and D, tube D into $\{g_i''\}$ (see Figure 23.7). This eliminates the intersections between $\{\Sigma_i\}$ and D, at the expense of adding more

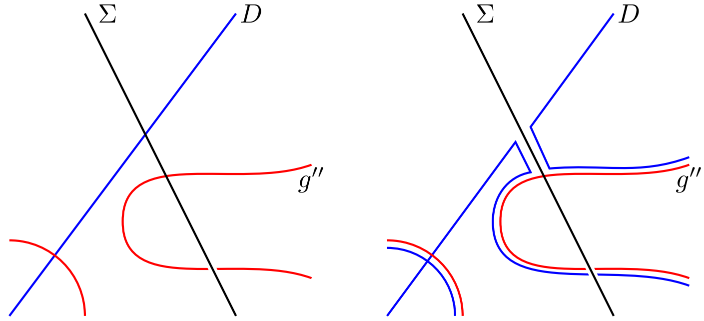


Figure 23.7. Resolving intersections between a surface Σ and a Whitney disc D pairing intersections between its caps and a transverse sphere.

intersections between D and $\{g_i''\}$ and new self-intersections of D, but we do not mind for the next step. Perform the above process for each Whitney disc pairing intersection points between a cap and $\{g_i''\}$.

Next, using the Whitney discs we just constructed, apply the geometric Casson lemma (which includes performing Whitney moves) to obtain $\{g_i'''\}$, regularly homotopic to $\{g_i''\}$, such that $\{g_i'''\}$ is disjoint from the collection of caps $\{C_\ell\}$, at the expense of adding self-intersections of $\{g_i'''\}$, and intersections, including self-intersections, among the caps. (Since $\{\Sigma_i\}$ did not intersect the Whitney discs we use in the argument, the $\{\Sigma_i\}$ are unchanged.) Now we have finally replaced the original set of immersed spheres $\{f_i\}$ by a union-of-spheres-like capped grope, namely $\{\Sigma_i\}$ equipped with the latest family of caps, with a geometrically transverse collection $\{g_i'''\}$ of immersed spheres. Since the geometric Casson lemma supplies a regular homotopy, we still have that $\lambda(C_\ell, C_k) = \mu(C_\ell) = 0$ for every pair of current caps C_ℓ and C_k, as well as $\mu(g_i''') = 0$ for all i, k, ℓ.

Now we will promote the caps in the present union-of-spheres-like capped grope to capped surfaces using a very similar process to the above. Since there is no material difference from what we have already described in detail, we run through the argument a little faster this time. Let $\{\alpha_m\}$ be the collection of curves on $\{\Sigma_i\}$ bounding the present collection of caps. Since the caps have zero intersection and self-intersection numbers, as before tube along Whitney circles to find surfaces $\{\sigma_m\}$ attached to the curves $\{\alpha_m\}$ (Figure 23.4). Use Whitney discs and meridional discs to cap each σ_m. Push down and tube into parallel push-offs of $\{g_i'''\}$ to eliminate intersections between the newest set of caps and $\{\Sigma_i\} \cup \{\sigma_m\}$. This results in a capped grope of height 2, but the caps intersect $\{g_i'''\}$.

Our strategy to address this is also similar to before. For each intersection point between $\{g_i'''\}$ and a cap, push the intersection down into the surface $\{\Sigma_i\}$ (each original intersection now yields four points of intersection with $\{\Sigma_i\}$) and tube $\{g_i'''\}$ into parallel copies of the pre-pushing $\{g_i'''\}$, as earlier in Figure 23.5. To avoid a comical proliferation of symbols, we still refer to the new spheres as $\{g_i'''\}$, as in Figure 23.6. For intersection points among the caps, including self-intersections, push the intersection down to $\{\Sigma_i\}$ and tube into parallel copies of $\{g_i'''\}$. Since all intersections now appear in pairs,

$$\mu(g_i''') = \lambda(g_i''', C_p') = \lambda(C_p', C_q') = \mu(C_p') = 0$$

for every i, p, q, where $\{C'_p\}$ is the latest collection of caps for $\{\Sigma_i\} \cup \{\sigma_m\}$. Now we have Whitney discs pairing the intersection points between $\{g'''_i\}$ and the caps. For each intersection between such a Whitney disc D' and $\{\Sigma_i\} \cup \{\sigma_m\}$, push down and tube D' into $\{g'''_i\}$, as in Figure 23.7. Use the resulting Whitney discs to apply the geometric Casson lemma to acquire spheres $\{\overline{g}_i\}$ geometrically transverse to the (unchanged) surfaces $\{\Sigma_i\}$ and disjoint from the $\{\sigma_m\}$ and the caps. The caps change by a regular homotopy in this process.

Apply grope height raising (Proposition 17.3) to replace $\{f_i\}$ by an arbitrarily tall union-of-spheres-like capped grope F', geometrically transverse to the immersed spheres $\{\overline{g}_i\}$. Tube each intersection within $\{\overline{g}_i\}$ into parallel copies of F'. This replaces $\{\overline{g}_i\}$ by a union-of-spheres-like capped grope G'. Since $\{\overline{g}_i\}$ has algebraically cancelling intersections, the tubing does not change the homotopy class.

In summary, we now have that the grope F' represents the basis of P, the grope G' represents the basis of P^*, and the gropes F' and G' are geometrically transverse, apart from the caps. To complete the proof, we need only remove intersections amongst the caps.

Enumerate the top stage surfaces of $F' \cup G'$. Iteratively, contract the ith surface and push off the caps of the surfaces numbered greater than i, as in the proof of the sequential contraction lemma (Lemma 17.7). This results in a pairwise disjoint collection of sphere-like capped gropes $\{F^c_i\}$, coming from F' and a pairwise disjoint collection of sphere-like capped gropes $\{G^c_i\}$, coming from G'. By construction, the collections $\{F^c_i\}$ and $\{G^c_i\}$ are geometrically transverse, and the collection of all caps is mutually disjoint. Moreover, by using grope height raising and contraction if necessary, we can arrange for each F^c_i and G^c_i to have arbitrarily chosen large heights. This completes the proof of Lemma 23.30. $\square$

We return to the proof of Proposition 23.29. By Lemma 23.30, we have arranged for the surgery kernel to be represented by geometrically transverse collections of sphere-like capped gropes $\{F^c_i\}$ and $\{G^c_i\}$, with arbitrary heights and with mutually disjoint caps. Let

$$W := \bigcup \{F^c_i\} \cup \bigcup \{G^c_i\}.$$

As an illustration, the simplest possible such W, corresponding to a single pair of geometrically transverse sphere-like capped gropes of height one, genus one in the surface stage, and mutually disjoint caps with a single self-intersection point each, is shown in Figure 23.8. We give both a schematic picture and a Kirby diagram for W. The bottom picture can be obtained from the middle picture using elementary Kirby moves, including replacing a dotted circle by a circle decorated with a zero framing. The bottom picture is a surgery diagram for ∂W. In particular, ∂W is the result of 0-framed Dehn surgery along each component of an element of $\mathcal{L}_2$.

Observe that the fundamental group of W is freely generated by the double point loops of $\{F^c_i\}$ and $\{G^c_i\}$, namely meridians of the link L. This determines an isomorphism to the free group $\Psi \colon \pi_1(W) \to F_n$, where n is the total number of double point loops in $\{F^c_i\} \cup \{G^c_i\}$. As already noted, $\partial W = M_L$ for some an n-component link $L \in \mathcal{L}_2$. We observe that $\Psi|_{\partial W} = \psi \colon \pi_1(\partial W) \to F_n$ is a homomorphism with perfect kernel exhibiting that (L, ψ) is a good boundary link. We also note that (W, Ψ) is a null bordism for (M_L, ψ) over F_n. Since W is smooth and spin [**GS99**, Corollary 5.7.2], apply Lemma 23.19 to obtain a degree one normal map

$$g \colon (W, \partial W) \to (\natural_{i=1}^n S^1 \times D^3, \#_{i=1}^n S^1 \times S^2)$$

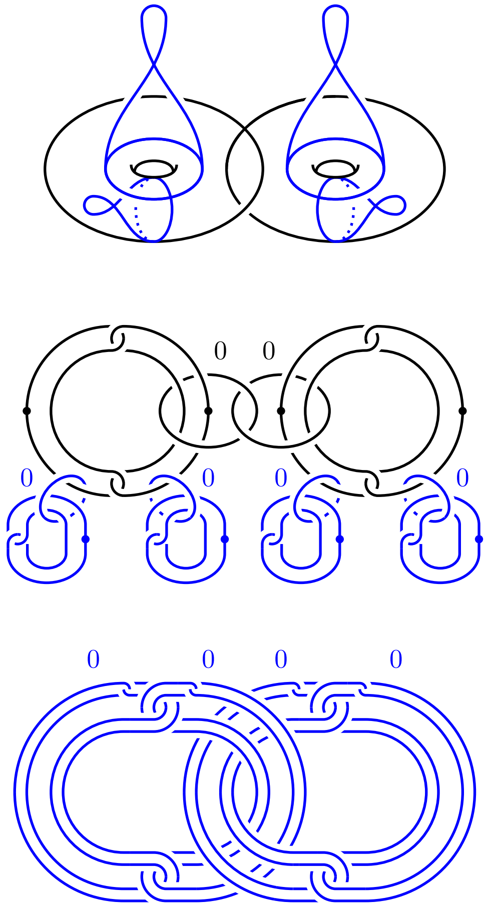


Figure 23.8. The simplest possible W, corresponding to a single pair of geometrically transverse sphere-like capped gropes of height one, genus one in the surface stage, and mutually disjoint caps with a single self-intersection point each. Top: A schematic diagram for the 2-dimensional spine of W. Middle: A Kirby diagram for W. Bottom: A Dehn surgery diagram for ∂W.

inducing the given maps (Ψ, ψ) on the fundamental groups. As in the proof of Proposition 23.14, we may take appropriate connected sums of W and the E_8-manifold to modify g to have vanishing surgery obstruction.

By hypothesis, the link L is freely topologically slice in D^4. In other words, there exists a 4-manifold W' with boundary $\partial W' = M_L$, namely the exterior of these slice discs, with a degree one normal map

$$g' \colon (W', \partial W') \to (\natural_{i=1}^n S^1 \times D^3, \#_{i=1}^n S^1 \times S^2)$$

as in the proof of Lemma 23.23 (see also [**FQ90**, 11.6C(1)]), agreeing with ψ on the boundary and such that the map $W' \to \natural_{i=1}^n S^1 \times D^3$ is a homotopy equivalence. We also saw in the proof of Lemma 23.23 that the collection of degree one normal maps relative to the boundary, with target $(Y, \partial Y) := (\natural_{i=1}^n S^1 \times D^3, \#_{i=1}^n S^1 \times S^2)$, is the set of homotopy classes

$$[(Y, \partial Y), G/TOP] = H^2(Y, \partial Y; \mathbb{Z}/2\mathbb{Z}) \oplus H^4(Y, \partial Y; \mathbb{Z}) \cong 0 \oplus \mathbb{Z},$$

where the element in the latter $\mathbb{Z}$ is given by the surgery obstruction. Since the maps g and g' agree on the boundary and both have vanishing surgery obstructions, they are normally bordant. In other words, the map g' is a solution for the surgery problem given by g, with the kernel associated with the normal bordism the given stable lagrangian. Thus, by [**FK16**, Lemma 7.3], it follows that

$$\phi' \colon (M', \partial M') := ((M \smallsetminus W) \cup_{\partial W} W', \partial M) \to (X, \partial X)$$

is a solution for $\phi\colon (M, \partial M) \to (X, \partial X)$, as needed. □

Other universal link slicing reformulations of the surgery conjecture, in the vein of Propositions 23.25 and 23.29, have been developed by Freedman and Krushkal in [**FK16**, Section 7; **FK20**, pp. 7–8]. Inspired by such reformulations, many freely topologically slice Whitehead doubles were constructed in [**Fre85**, **Fre88**, **FT95b**]. The most general of these, due to Freedman and Teichner [**FT95b**], says that Whitehead doubles of *homotopically trivial*$^+$ links L are freely topologically slice. Here a link L is *homotopically trivial*$^+$ if $L \cup L_i^+$ is homotopically trivial for every i, where L_i^+ is the 0-framed parallel copy of the ith-component L_i of L. In the proof of [**FT95b**, Theorem 3.1], the existence of immersed discs bounded by $L \cup L_i^+$ with certain disjointness properties was crucial, and it seems to be very difficult to generalise the argument for untwisted Whitehead doubles of L when L is not link homotopic to the unlink. Unfortunately, this is the case for GBRs, as these links have nontrivial, non-repeating Milnor invariants, by the Bing doubling formula for Milnor invariants given in [**Coc90**, Chapter 8].

In particular, it has been conjectured that the Whitehead double of a link L is freely topologically slice if and only if L is homotopically trivial (see [**CFT09**, Conjecture 1.1; **CP16**, Conjecture 1.1]). Thus, for example, it is a very interesting question whether the Whitehead double of the Borromean rings, which is a good boundary link by Remark 23.13, is freely topologically slice in D^4. We have the following conjecture and equivalence.

Conjecture 23.31 (Conjecture GBL). *Every good boundary link is freely topologically slice.*

Proposition 23.32. *Conjecture GBL is equivalent to the surgery conjecture for all groups.*

Proof. The surgery conjecture implies Conjecture GBL by Proposition 23.21. For the other direction, there are two routes, as shown in Figure 23.3. One way is to note that each element of $\mathcal{L}_2$ is a good boundary link, and invoke Proposition 23.29. The other option is to note that every element of $\mathcal{L}_1$ is a good boundary link, and invoke Propositions 23.25 and 23.26. □

Other constructions of freely topologically slice good boundary links are given in [**Fre93**, **CKP20**]. The construction in [**CKP20**] subsumes all previous methods for freely topologically slicing good boundary links with two or more components (including [**Fre85**, **Fre88**, **Fre93**, **FT95b**]). We briefly mention the main result of [**CKP20**]. Recall that every good boundary link has a good basis $\{a_i, b_i\}_{i=1}^g$ (Proposition 23.12). A good basis $\{a_i, b_i\}_{i=1}^g$ is *homotopically trivial*$^+$ if $J \cup a_i$ and $J \cup b_i$ are homotopically trivial for all i where J is the union $\bigcup_{i=1}^g b_i'$ of a parallel copy b_i' of b_i such that $\operatorname{lk}(a_i, b_i') = 0$. The main result of [**CKP20**] is the following theorem.

Theorem 23.33. *Every good boundary link L with a boundary link Seifert surface admitting a homotopically trivial*$^+$ *good basis is freely topologically slice.*

23.2.4. The A-B slice problem. In both of our classes of universal links, $\mathcal{L}_1$ and $\mathcal{L}_2$, we perform a final step of ramified Whitehead doubling. This doubling makes it rather difficult to obstruct these links from being slice directly. As a simple example of this phenomenon, note that such Whitehead doubles are all boundary links and thus have trivial Milnor invariants. In [**Fre86a**], Freedman introduced the notion of *A-B slice links*. Their definition is somewhat complicated, so we first describe some properties. If L is a GBR, it is known that L is A-B slice if and only if every ramified Whitehead double on all components of L is freely topologically slice [**Fre86b**, **Fre86a**]. We have the following conjecture and equivalence.

Conjecture 23.34 (Conjecture AB). *Every GBR is A-B slice.*

Proposition 23.35. *Conjecture AB is equivalent to the surgery conjecture for all groups.*

Proof. Assume Conjecture AB. Then as mentioned above, the ramified Whitehead doubles of GBRs are freely slice [**Fre86b**, **Fre86a**]. In particular, the elements of $\mathcal{L}_2$ are freely slice by the definition of $\mathcal{L}_2$. Then the surgery conjecture follows by Proposition 23.29.

Now assume the surgery conjecture. Then good boundary links are freely slice by Proposition 23.21. Then in particular ramified Whitehead doubles of GBRs are freely slice (Remark 23.13). Again using the result of [**Fre86b**, **Fre86a**], we see that all GBRs are A-B slice. □

Thus, in particular, if the Borromean rings are not A-B slice, the surgery conjecture is false. Of course, this would imply that not all groups are good.

We now give the definition of A-B slice links. Define a *decomposition* of D^4 to be a pair $A, B \subseteq D^4$ of smooth, compact, codimension zero submanifolds with corners satisfying the following conditions.

(1) $A \cup B = D^4$.
(2) $\partial A = \partial_- A \cup \partial_+ A$, where $\partial_+ A = \partial A \cap \partial D^4$ and the corner set $\partial_- A \cap \partial_+ A$ is the standard torus $\partial D^2 \times \partial D^2 \subseteq \partial D^4$.
(3) $\partial B = \partial_- B \cup \partial_+ B$, where $\partial_+ B = \partial B \cap \partial D^4$ and the corner set $\partial_- B \cap \partial_+ B = \partial_- A \cap \partial_+ A$.
(4) $A \cap B = \partial_- A = \partial_- B$.

We can interpret a decomposition of D^4 to be an extension in D^4 of the standard genus one Heegaard decomposition of S^3. This use of the word decomposition should not be confused with the decompositions studied in Part I and put to use in Part IV.

Let L be a link in S^3 with k components. Let $D(L)$ be the $2k$-component link obtained by adding to L a 0-framed parallel for each component of L. We say that L is *A-B slice* if there exist k decompositions of D^4, $(A_1, B_1), \ldots, (A_k, B_k)$, and $2k$ self-homeomorphisms of D^4, $\alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_k$, such that the entire collection $\{\alpha_1(A_1), \ldots, \alpha_k(A_k), \beta_1(B_1), \ldots, \beta_k(B_k)\}$ is pairwise disjoint and satisfies that $\alpha_i(\partial_+ A_i)$ is a tubular neighbourhood of the ith component of L and $\beta_i(\partial_+ B_i)$ is a tubular neighbourhood of the parallel copy of the ith component of L, for all $1 \leq i \leq k$. In particular, the union $\bigcup_{i=1}^k (\alpha_i(\partial_+ A_i) \cup \beta_i(\partial_+ B_i))$ is the link $D(L)$.

All GBRs are *link-homotopy A-B slice* [**FK16**, Theorem 1] and some GBRs, which also form a universal family of links, are known to be *(link-homotopy)$^+$ A-B slice* [**FK17a**, Theorem 1]. We stress that the 'link-homotopy' and '(link-homotopy)$^+$' qualifiers are not intended to apply to the GBRs themselves but rather are applied to certain links obtained in the description of the A-B slice decomposition. These results imply that (after stabilisation and band sum) the GBRs in question have vanishing Milnor invariants with at most two repeating indices, indicating the resilience of the A-B slice problem to known link invariants.

This completes our discussion of the surgery conjecture.

23.3. The s-cobordism conjecture

We return to discussing the disc embedding conjecture, explaining the remaining implications shown in Figure 23.1. In addition to the surgery conjecture, the disc embedding conjecture implies the *s-cobordism conjecture*, which we now state.

CONJECTURE 23.36 (s-cobordism conjecture). *Every compact, topological 5-dimensional s-cobordism $(W; M_0, M_1)$, relative to the boundary, is homeomorphic to the product $M_0 \times [0,1]$, extending the given product structure on the boundary $\partial M_0 = \partial M_1$.*

Indeed, as we now establish, the disc embedding conjecture is equivalent to the combination of the surgery and s-cobordism conjectures.

PROPOSITION 23.37. *The disc embedding conjecture is equivalent to the logical union of the surgery conjecture and the s-cobordism conjecture.*

PROOF. See Chapters 20 and 22 for the proofs that the disc embedding conjecture implies the s-cobordism and surgery conjectures, in the case of closed manifolds, as always using the immersion lemma, topological transversality, and the existence of normal bundles for locally flat submanifolds of a topological 4-manifold. The statement for compact manifolds can be proven similarly.

For the converse, we will show that the union of the surgery and s-cobordism conjectures implies that each element of $\mathcal{L}_1$ is topologically slice in D^4, where the slice discs for $L \smallsetminus m(L)$ are standard. In other words, we will show Conjecture $\mathcal{L}_1$, which will complete the proof by Proposition 23.5.

Let $L \in \mathcal{L}_1$ be an $(n+1)$-component link. Recall from Remark 23.13 that all links in $\mathcal{L}_1$ are good boundary links. Thus, by Proposition 23.21, L is freely topologically slice. Choose such a collection of free slice discs and let X be the exterior of the slice discs for the unlink $L \smallsetminus m(L)$. By Lemma 23.23, there is an s-cobordism (relative to the boundary) from X to $\natural_{i=1}^n S^1 \times D^3$. By the s-cobordism conjecture, there is a homeomorphism (relative to the boundary) $f\colon X \to \natural_{i=1}^n S^1 \times D^3$. But the slice discs for L in D^4 determine a slice disc in X for $m(L) \subseteq \partial X$. Using f, this means $m(L) \subseteq \#_{i=1}^n S^1 \times S^2$ is slice in $\natural_{i=1}^n S^1 \times D^3$. In other words, $m(L)$ is slice in D^4 away from a collection of standard slice discs for the other components of L, as claimed. ☐

23.3.1. Round handles. As noted earlier, the surgery conjecture would imply that all GBRs are A-B slice. A corresponding statement implied by the combination of the surgery and s-cobordism conjectures was proposed and investigated in [**FK16**, Section 5.1] (see also [**KPT17**]).

A *round handle* is a copy of $D^1 \times D^2 \times S^1$, which is attached to the boundary of a 4-manifold along $S^0 \times D^2 \times S^1$. Given an m-component link $L \subseteq S^3$ we construct a 4-manifold $R(L)$ by attaching m round handles to D^4 as follows. For the ith component L_i of L, let λ_i denote a 0-framed longitude and μ_i denote a meridian, chosen so that $\ell k(\lambda_i, \mu_i) = 0$. To attach the ith round handle, identify $\{-1\} \times S^1 \times D^2$ to λ_i and $\{+1\} \times S^1 \times D^2$ to μ_i, using the trivial framing in both cases. The resulting 4-manifold contains the link L in the boundary. A link L is said to be *round handle slice* if $L \subseteq \partial R(L)$ is slice in $R(L)$, that is, if L is the boundary of a collection of locally flat pairwise disjoint embedded discs in $R(L)$. We have the following conjecture and implication.

CONJECTURE 23.38 (Conjecture RH). *Every link with zero pairwise linking numbers is round handle slice.*

PROPOSITION 23.39 ([**FK16**, Section 5.1] (see also [**KPT17**])). *The surgery and s-cobordism conjectures together imply Conjecture RH.*

Unlike the A-B slice problem, it is not currently known whether the round handle problem has a solution up to link homotopy. Thus, non-repeating Milnor invariants might yet provide an effective obstruction for this problem.

Part IV

Skyscrapers are standard

In Part II we saw how to begin with the hypotheses of the disc embedding theorem and produce a mutually disjoint collection of skyscrapers whose attaching regions coincide with the framed boundaries of Whitney discs pairing up all the intersection points among the initial collection of immersed discs. In this, the final part, we complete the proof of the disc embedding theorem, as explained in Section 18.4, by showing that every skyscraper is homeomorphic to a 2-handle relative to its attaching region. The proof is given in detail in Chapter 28, and uses several facts from both Parts I and II. The argument is rather complicated, both mathematically and in terms of notation. We set the stage for the proof by first pointing out exactly which properties of skyscrapers are necessary, in Chapter 24. In Chapter 25 we prove the *collar adding lemma*, using techniques similar to our work with the Alexander gored ball in Chapter 5, which states that a skyscraper with an additional boundary collar glued on is homeomorphic to the standard ball $D^2 \times D^2$. This lemma is the last remaining ingredient needed for the proof in Chapter 28. We collect the key ingredients from Parts I and II in Chapter 26 for the convenience of the reader. We give an outline of the proof that skyscrapers are homeomorphic to 2-handles in Chapter 27, before filling in all the details in Chapter 28.

CHAPTER 24

Replicable rooms and boundary shrinkable skyscrapers

Stefan Behrens and Mark Powell

At this point, we have expended considerable effort to progress from the hypotheses of the disc embedding theorem to certain infinite iterated objects, namely skyscrapers. Remarkably, in the remainder of the proof, only certain key properties of skyscrapers will be utilised, rather than the specific details of their architecture. In this chapter we define a class of objects generalising towers and skyscrapers, distilling the properties that we need going forward. In particular, this should clarify the reasoning behind constructing skyscrapers instead of Casson handles. The construction of Casson handles [**Cas86**, **Fre82a**, **GS84**] is somewhat easier than that of skyscrapers. However, the shrinking in Chapter 28 is substantially easier for skyscrapers. In particular, the upcoming argument uses that skyscrapers can be explicitly understood in relation to the standard 2-handle $D^2 \times D^2$ in terms of Kirby diagrams, that they contain compatibly embedded copies of themselves with higher levels squeezed into small balls, and that they have nice shrinking properties on the boundary. This last property is not enjoyed by Casson handles. While Freedman was able to work around this problem, his argument is harder.

Additionally, the reader might also see the opportunity to develop other constructions satisfying the requirements of the present chapter. Such a construction would ideally help answer open questions regarding good groups or topological surgery. As long as the construction satisfies the requirements we lay out in this chapter, the arguments of Chapter 28 will still apply.

This chapter is not logically required for the rest of the proof, and could be skipped in the first reading. However, knowing the precise properties of skyscrapers needed in the sequel might be useful for a deeper understanding.

Definition 24.1. A *room* is a triple (R, K_-, ϕ) where R is a smooth, compact, connected 4-manifold with corners, K_- is a framed knot in ∂R, and after smoothing corners, R is diffeomorphic via ϕ to a boundary connected sum $\natural^k S^1 \times D^3$, for some $k \geq 1$, represented by a Kirby diagram consisting of a k-component dotted unlink L_R in S^3. A room (R, K_-, ϕ) must additionally satisfy the following conditions.

1. The framed knot K_- is carried by ϕ to a framed knot in $S^3 \smallsetminus L_R$, which is a 0-framed unknot in S^3.
2. The link L_R is null-homotopic in $S^3 \smallsetminus \phi(\nu K_-) \cong D^2 \times S^1$.

The closed tubular neighbourhood $\nu K_- \subseteq \partial R$ defined by the framing is called the *attaching region* of the room (R, K_-, ϕ) and is denoted by $\partial_- R$. Note that, via ϕ, the boundary of R is diffeomorphic to the result $\#^k S^1 \times S^2$ of 0-framed Dehn surgery on S^3 along the components of L_R. Thus we have a decomposition of ∂R as $\partial_- R \cup \partial_+ R \cup \overline{\partial} R$, where $\partial_+ R$ corresponds via ϕ to

$$\operatorname{cl}\left(S^3 \smallsetminus (\nu L_R \cup \phi(\nu K_-))\right) \cong \operatorname{cl}\left(D^2 \times S^1 \smallsetminus \nu L_R\right),$$

and is called the *walls* of the room (R, K_-, ϕ), while $\overline{\partial} R$ corresponds to the collection of surgery solid tori, and is called the *ceiling* of the room (R, K_-, ϕ).

The corners of R are precisely the components of the boundary of the walls $\partial_+ R$, or equivalently the union of the boundary of the attaching region $\partial_- R$ and the boundary of the ceiling $\overline{\partial} R$.

REMARK 24.2. In Definition 24.1, let J be a fixed collection of 0-framed (with respect to S^3) meridians of L_R, which freely generate $\pi_1(\natural^k S^1 \times D^3)$. The set J corresponds via ϕ to a collection of k framed circles in ∂R. Note that the ceiling of a room (R, K_-, ϕ) is isotopic, with its induced framing as surgery tori, to 0-framed neighbourhoods for the meridians J. Compare with Section 13.4.

REMARK 24.3. In Definition 24.1, the condition that L_R be null-homotopic does not refer to homotopy through *link maps*. In that theory, components are required to have disjoint images throughout a homotopy. It is null-homotopic in the simpler sense that L_R is the image of a null-homotopic map from a disjoint union of circles to $D^2 \times S^1$. In other words, each component of L_R has winding number zero.

In our standard terminology, any surface, disc, or cap block, disc-like grope, disc-like capped grope, disc-like tower, of disc-like capped tower, along with their attaching regions, and usual corners, is a room. The map ϕ is provided by the Kirby diagrams in Chapter 13. The ceiling corresponds to the tip region and the walls correspond to the vertical boundary.

DEFINITION 24.4. A *level* is a disjoint union of finitely many rooms. The attaching region, the walls, and the ceiling of a level consist of the (disjoint) union of the attaching regions, walls, and ceilings of these rooms respectively.

REMARK 24.5. We avoid using the word floor since by an unfortunate quirk in the english language a floor might correspond either to an attaching region or to a level in our architectural analogy.

In the previous terminology, every grope, capped grope, tower, or capped tower (along with its attaching region, and its usual corners) is a level. We can stack levels to produce buildings. We give an inductive definition.

DEFINITION 24.6. A *building* with one level is simply a level. A *building* $\mathcal{B}_{n+1}$ with $n+1$ levels is obtained by attaching the ceiling of a building $\mathcal{B}_n$ with n levels to the attaching region of a building $\mathcal{B}_1$ with one level, by matching the framings. This requires that $\mathcal{B}_1$ have as many components as the ceiling of $\mathcal{B}_n$. The corner points that are glued together become ordinary boundary points. The *attaching region* $\partial_- \mathcal{B}_{n+1}$ of $\mathcal{B}_{n+1}$ is the attaching region of $\mathcal{B}_n$ while the *ceiling* $\overline{\partial} \mathcal{B}_{n+1}$ of $\mathcal{B}_{n+1}$ is the ceiling of $\mathcal{B}_1$. The *walls* $\partial_+ \mathcal{B}_{n+1}$ of $\mathcal{B}_{n+1}$ consist of the union of the walls of $\mathcal{B}_n$ and the walls of $\mathcal{B}_1$.

As before for the constructions in Chapter 12 (see Remark 12.7), the decomposition into levels is part of the data of a building.

Using the techniques from Chapter 13, we see that every connected building $\mathcal{B}$ is also diffeomorphic to $H \smallsetminus (\coprod^{\ell} D^2 \times \mathring{D}^2)$ for some ℓ, where $H = D^2 \times D^2$ and each copy of $S^1 \times \mathring{D}^2$ lies in $D^2 \times S^1 \subseteq \partial H$. Utilising handle slides, as in Section 13.6, we may produce a link L_B in S^3 for $\mathcal{B}$ from the diagrams for the constituent parts. Note that L_B is also an unlink in S^3 consisting of dotted circles, and is null-homotopic in $S^3 \smallsetminus \partial_- \mathcal{B} \cong D^2 \times S^1$. In our standard terminology, any capped grope, 1-storey tower, or 1-storey capped tower is a building with one level. Every n-storey tower or n-storey capped tower is a building with n levels. Indeed, any n-storey capped tower is a building with k levels for each $k \leq n+1$, depending on our choice of constituent levels.

DEFINITION 24.7. An *infinite building* $\mathcal{B}_\infty$ is a building with countably infinitely many levels. The *attaching region* $\partial_-\mathcal{B}_\infty$ of $\mathcal{B}_\infty$ is the attaching region of any of its finite truncations, while the *walls* $\partial_+\mathcal{B}_\infty$ of $\mathcal{B}_\infty$ consist of the union of the walls of the finite truncations. Note that $\partial\mathcal{B}_\infty = \partial_-\mathcal{B}_\infty \cup \partial_+\mathcal{B}_\infty$.

An *infinite compactified building* $\widehat{\mathcal{B}}_\infty$ is the endpoint compactification of an infinite building $\mathcal{B}_\infty$. The *attaching region* $\partial_-\widehat{\mathcal{B}}_\infty$ of $\widehat{\mathcal{B}}_\infty$ is the attaching region of $\mathcal{B}_\infty$.

The ends of the walls of the infinite building $\mathcal{B}_\infty$ are in natural bijective correspondence with the ends of $\partial_+\mathcal{B}_\infty$ (cf. Remark 12.19). The *walls* $\partial_+\widehat{\mathcal{B}}_\infty$ of $\widehat{\mathcal{B}}_\infty$ are by definition the endpoint compactification of $\partial_+\mathcal{B}_\infty$. We also write $\partial\widehat{\mathcal{B}}_\infty := \partial_-\widehat{\mathcal{B}}_\infty \cup \partial_+\widehat{\mathcal{B}}_\infty$, which also coincides with the endpoint compactification of $\partial\mathcal{B}_\infty$.

In the standard terminology every infinite tower, along with its attaching region, is an infinite building. Similarly every infinite compactified tower, along with its attaching region, is an infinite compactified building.

By construction, a connected building is produced from $D^2 \times D^2$ by removing several disjoint copies of $D^2 \times \mathring{D}^2$, that is, thickened discs, corresponding to the 1-handles. Moreover, the thickened discs for the successive truncations of an infinite building are nested. Thus we define a decomposition $\mathcal{D}_{\mathcal{B}_\infty}$, corresponding to any infinite connected building $\mathcal{B}_\infty$, given by the infinite intersection of the thickened discs corresponding to the finite truncations. In other words, each element of $\mathcal{D}_{\mathcal{B}_\infty}$ is a connected component of the infinite intersection of the thickened discs corresponding to the finite truncations of $\mathcal{B}_\infty$. We see then, as in Chapter 14, that the infinite connected building $\mathcal{B}_\infty$ is the complement

$$\mathcal{B}_\infty \cong D^2 \times D^2 \smallsetminus \bigcup_{\Delta_i \in \mathcal{D}_{\mathcal{B}_\infty}} \Delta_i,$$

while the corresponding infinite compactified building $\widehat{\mathcal{B}}_\infty$ is the associated decomposition space

$$\widehat{\mathcal{B}}_\infty \cong D^2 \times D^2/\mathcal{D}_{\mathcal{B}_\infty}.$$

We define the corresponding decomposition on the boundary, by setting

$$\partial\mathcal{D}_{\mathcal{B}_\infty} := \{\Delta_i \cap \partial_+\mathcal{B}_\infty \mid \Delta_i \in \mathcal{D}_{\mathcal{B}_\infty}\}$$

so that, similarly to Chapter 14,

$$\partial_+\mathcal{B}_\infty \cong D^2 \times S^1 \smallsetminus \bigcup_{\Delta_i \in \partial\mathcal{D}_{\mathcal{B}_\infty}} \Delta_i,$$

and

$$\partial_+\widehat{\mathcal{B}}_\infty \cong D^2 \times S^1/\partial\mathcal{D}_{\mathcal{B}_\infty}.$$

Now we give two further restrictions on infinite compactified buildings.

DEFINITION 24.8. An infinite compactified building is said to be *boundary shrinkable* if for each connected component $\widehat{\mathcal{B}}_\infty$, the decomposition $\partial\mathcal{D}_{\mathcal{B}_\infty}$ is shrinkable, and moreover there is a sequence of homeomorphisms $h_i\colon D^2 \times S^1 \to D^2 \times S^1$, $i = 1, 2, \ldots$, with each h_i isotopic to the identity via an isotopy supported in the ith stage of the defining sequence for $\partial\mathcal{D}_{\mathcal{B}_\infty}$, such that

$$\lim_{m\to\infty} h_m \circ h_{m-1} \circ \cdots \circ h_1$$

exists and has the same inverse sets as the quotient map $D^2 \times S^1 \to D^2 \times S^1/\partial\mathcal{D}_{\mathcal{B}_\infty}$.

By Chapter 8 and Definition 12.21, the skyscrapers from Part II are boundary shrinkable. More precisely, any disc-like infinite compactified tower $\widehat{\mathcal{T}}_\infty$ such that $\sum_{i=1}^\infty N_j/2^j$ diverges, where N_j is the number of surface stages in the jth storey of $\widehat{\mathcal{T}}_\infty$, is boundary shrinkable.

Definition 24.9. Let $\mathcal{E}$ be a collection of rooms. An *infinite compactified $\mathcal{E}$-building* is an infinite compactified building all of whose constituent rooms lie in the class $\mathcal{E}$.

Definition 24.10. A class of rooms $\mathcal{E}$ is said to be *replicable* if for every room $R \in \mathcal{E}$ and for each open set $U \subseteq R \smallsetminus \partial R$ there is an embedding of a boundary shrinkable infinite compactified $\mathcal{E}$-building $\widehat{\mathcal{B}}_\infty$ with the same attaching region as R and such that all the levels of $\widehat{\mathcal{B}}_\infty$ above the first level lie inside U. Moreover, $\partial_+\widehat{\mathcal{B}}_\infty$ and $\partial_+ R$ coincide in a neighbourhood of $\partial_- R$.

The work of Chapter 18 can be rephrased as the following.

Theorem 24.11. *The class of disc-like* 2*-storey towers with at least four surface stages in the first storey forms a replicable class of rooms.*

We are finally ready to define the generalisation of skyscrapers. To underscore the analogy with skyscrapers, we refer to them by the same name.

Definition 24.12. An infinite compactified building $\widehat{\mathcal{B}}_\infty$ is said to be a *skyscraper* if it is boundary shrinkable and an infinite compactified $\mathcal{E}$-building for some replicable class of rooms $\mathcal{E}$.

We immediately see from the definition that Definition 24.12 is a generalisation of the previous notion of skyscrapers. While we will return to talking about skyscrapers as defined in Part II starting from the next chapter, the reader will observe that we only ever use the properties described in this chapter. The remainder of this book will prove the following theorem.

Theorem 27.1. *For every skyscraper $\widehat{\mathcal{S}}$ there is a homeomorphism of pairs*

$$(F, \partial F)\colon \big(\widehat{\mathcal{S}}, \partial_-\widehat{\mathcal{S}}\big) \cong \big(D^2 \times D^2, S^1 \times D^2\big)$$

that is a diffeomorphism on a collar of $\partial_-\widehat{\mathcal{S}}$, such that if $\Phi\colon S^1 \times D^2 \to \partial_-\widehat{\mathcal{S}}$ is the attaching region, then $\partial F \circ \Phi = \mathrm{Id}_{S^1 \times D^2}$.

In particular, this will show that any room lying within a replicable class contains a flat embedded disc whose framed boundary coincides with the attaching region of the room.

CHAPTER 25

The collar adding lemma

Daniel Kasprowski and Mark Powell

We prove the collar adding lemma, which generalises the fact explained in Chapter 5 (Remark 5.2) that the Alexander gored ball with an added collar is homeomorphic to the standard 3-ball. This is the last remaining ingredient needed for the conclusion of the proof of the disc embedding theorem in Chapter 28.

LEMMA 25.1 (Collar adding lemma). *Let $\widehat{\mathcal{S}}$ be a skyscraper. Then the space*

$$\widehat{\mathcal{S}} \cup_{\mathrm{Id}} \left(\partial\widehat{\mathcal{S}} \times [0,1]\right),$$

glued along $\partial\widehat{\mathcal{S}} \times \{0\}$, is homeomorphic to $D^2 \times D^2$.

Although we have seen the corresponding fact in the 3-dimensional case before, it is still rather remarkable: simply adding a collar nullifies all the potentially interesting topology within a skyscraper. Note that we may not assume that the skyscrapers we have built in the proof of the disc embedding theorem in Part II come equipped with such external collars in the ambient 4-manifold. This fact engenders an impressive amount of extra difficulty, as will become apparent in Chapter 28.

We saw in Chapter 14 that $\widehat{\mathcal{S}}$ is homeomorphic to the decomposition space $(D^2 \times D^2)/\mathcal{D}$ corresponding to a decomposition $\mathcal{D}$ and we have the associated quotient map

$$\pi_{\mathcal{D}} \colon D^2 \times D^2 \to (D^2 \times D^2)/\mathcal{D}.$$

The decomposition $\mathcal{D}$ of $D^2 \times D^2$ restricted to the boundary of $D^2 \times D^2$ induces a decomposition $\partial\mathcal{D}$ of $\partial(D^2 \times D^2)$. The set $\partial\mathcal{D}$ consists of the elements $\Delta \cap \partial(D^2 \times D^2)$ for $\Delta \in \mathcal{D}$. In our case, the elements of $\partial\mathcal{D}$ are connected, and indeed, $\partial\mathcal{D}$ is a mixed ramified Bing-Whitehead decomposition. The combination of the following two lemmas will give the proof of the collar adding lemma.

LEMMA 25.2. *We may assume that the decomposition $\partial\mathcal{D}$ of $\partial(D^2 \times D^2)$ consists only of singleton sets.*

PROOF. Within an interior collar $\partial(D^2 \times D^2) \times [0,1]$ of the boundary of $D^2 \times D^2$, such that $\partial(D^2 \times D^2) \times \{0\} = \partial(D^2 \times D^2)$, the decomposition $\mathcal{D}$ can be taken to be a product $\partial\mathcal{D} \times [0,1]$, that is, the product of a mixed ramified Bing-Whitehead decomposition with the interval. By Scholium 8.13, there is a sequence of ambient isotopies of $\partial(D^2 \times D^2)$ transforming the defining sequence of $\partial\mathcal{D}$ such that each decomposition element arising from the new defining sequence is a single point. Apply these isotopies to $\partial\mathcal{D}$ in the collar $\partial(D^2 \times D^2) \times [0,1]$, so that we get the new defining sequence on $\partial(D^2 \times D^2) \times \{0\} = \partial(D^2 \times D^2)$ and the original defining sequence on $\partial(D^2 \times D^2) \times \{1\}$. Since we have the identity map on the inner boundary of the collar, we can extend the homeomorphism to all of $D^2 \times D^2$. Now, on the boundary, the decomposition consists only of singleton sets as desired. As we saw in Proposition 4.7, the isotopy of the defining sequence changes the decomposition, but gives rise to a homeomorphic decomposition space. □

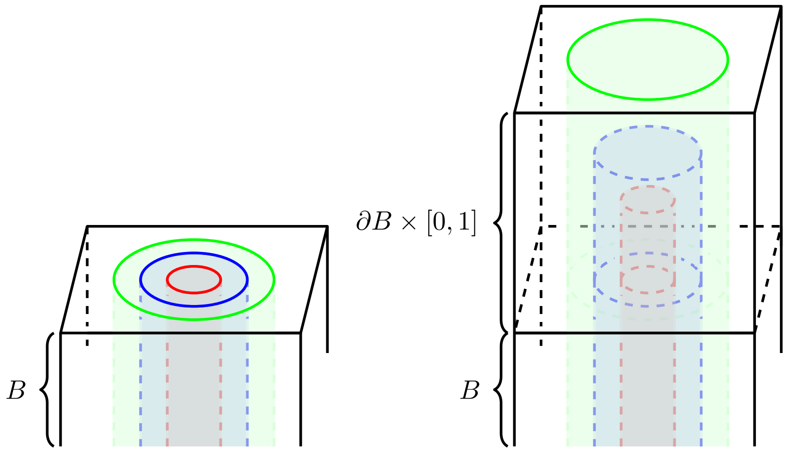


Figure 25.1. Adding a collar to a defining sequence. We depict the situation one dimension down. Left: The intersection of the nested solid cylinders $D^1 \times D^2$ (shown in red, blue, and green) close to the boundary of B. For each cylinder, the boundary $S^0 \times D^2$ lies in ∂B. We only show half of these boundaries, where the red disc is nested within the blue disc, which is nested within the green. On the right, we have attached a collar to B, and extended the solid cylinders within this collar. They remain nested, but now the entire red component is contained within the interior of the blue component, and the entire blue component lies within the interior of the green component. Take a product of the depicted situation with the interval to obtain a schematic in the correct dimension. In our situation, the nested elements of the defining sequence are unions of the form $S^1 \times D^2 \times [0,1]$, embedded as mixed ramified Bing-Whitehead doubles.

Once we make the restriction of Lemma 25.2, the map $\pi_{\mathcal{D}}$ is a homeomorphism when restricted to the boundary $\partial(D^2 \times D^2)$, and so we have

$$\begin{aligned} \widehat{\mathcal{S}} \cup_{\mathrm{Id}} \big(\partial\widehat{\mathcal{S}} \times [0,1]\big) &\cong (D^2 \times D^2)/\mathcal{D} \cup_{\mathrm{Id}} \big(S^3 \times [0,1]\big) \\ &= \big((D^2 \times D^2) \cup_{\mathrm{Id}} (S^3 \times [0,1])\big)/\mathcal{D}. \end{aligned} \tag{25.1}$$

Here the first isomorphism follows from Lemma 25.2, with $S^3 \times \{0\}$ identified with $\partial(D^2 \times D^2) = \partial\widehat{\mathcal{S}}$. In the final expression we are mildly abusing notation by considering $\mathcal{D}$ to be a decomposition of the expanded space $(D^2 \times D^2) \cup_{\mathrm{Id}} (S^3 \times [0,1])$.

Continuing with the same notation we have the following lemma.

Lemma 25.3. *The space* $\big((D^2 \times D^2) \cup_{\mathrm{Id}} (S^3 \times [0,1])\big)/\mathcal{D}$ *is homeomorphic to* $D^2 \times D^2$.

The collar adding lemma follows from combining (25.1) with Lemma 25.3.

Proof. Write $B := D^2 \times D^2$. The decomposition $\mathcal{D}$ of $B \cup_{\mathrm{Id}} (S^3 \times [0,1])$ is determined by a sequence D^i, $i = 1, 2, \ldots,$ where each D^i is a mutually disjoint union of copies of $D^2 \times D^2$, each embedded into D^{i-1} with the $S^1 \times D^2$ parts of their boundaries lying within $\partial B = S^3$, as shown in the left panel of Figure 25.1.

First we will arrange that for each i, each 2-handle in the collection D^i lies in the interior of D^{i-1}. Let T^i be the intersection $D^i \cap \partial B$. This is a mutually disjoint union of solid tori, one for each constituent 2-handle in D^i. Modify the defining sequence for $\mathcal{D}$ by taking the union

$$\widetilde{D}^i := D^i \cup_{\mathrm{Id}} \left(T^i \times \left[0, \frac{1}{i+1}\right]\right),$$

for each $i \geq 1$ Here each $D^i \subseteq B$ and $T^i \times [0, \frac{1}{i+1}] \subseteq \partial B \times [0,1] = S^3 \times [0,1]$ (see Figure 25.1). Let $\widetilde{\mathcal{D}}$ be the decomposition given by the defining sequence $\{\widetilde{D}^i\}$. Note that changing the defining sequence in this way does not alter the infinite intersection, so the decompositions are equal, $\widetilde{\mathcal{D}} = \mathcal{D}$. But, as desired, $\widetilde{D}^i$ is now contained in the interior of $\widetilde{D}^{i-1}$ for every $i \geq 2$.

The decomposition $\widetilde{\mathcal{D}}$ of the space $X := B \cup_{\mathrm{Id}} (S^3 \times [0,1])$ is upper semi-continuous by Proposition 4.10. Choose some metric $d_{X/\widetilde{\mathcal{D}}}$ on the quotient space $X/\widetilde{\mathcal{D}}$, which is metrisable by Corollary 4.13. It is straightforward to shrink the decomposition $\widetilde{\mathcal{D}}$ using the definition (Definition 4.16) (see also Theorems 4.1 and 4.18) as follows. Recall that given $\varepsilon > 0$, we must find a homeomorphism $h \colon X \to X$ that satisfies the following conditions.

(i) For all $x \in X$, we have that $d_{X/\widetilde{\mathcal{D}}}\big(\pi_{\widetilde{\mathcal{D}}}(x), \pi_{\widetilde{\mathcal{D}}} \circ h(x)\big) < \varepsilon$.

(ii) For all $y \in X/\widetilde{\mathcal{D}}$, we have that $\operatorname{diam}_X h(\pi_{\widetilde{\mathcal{D}}}^{-1}(y)) < \varepsilon$.

First we will show that given $\varepsilon > 0$, there exists $s > 0$ such that any homeomorphism which is the identity outside $\widetilde{D}^s$ only moves points by a distance less than ε in the metric of the target $X/\widetilde{\mathcal{D}}$. Let U be the union of ε-ball neighbourhoods of each point in the singular image of $\pi_{\widetilde{\mathcal{D}}}$, that is the collection of points in $X/\widetilde{\mathcal{D}}$ arising as the image of elements of $\widetilde{\mathcal{D}}$. We claim there is an s large enough so that $\widetilde{D}^s \subseteq \pi_{\widetilde{\mathcal{D}}}^{-1}(U)$. Suppose not. Then there is an infinite nested sequence $\{B_i\}_{i \geq 1}$, with each B_i a component of $\widetilde{D}_i$, and $B_i \not\subseteq \pi_{\widetilde{\mathcal{D}}}^{-1}(U)$. For each i, choose $x_i \in B_i \smallsetminus \pi_{\widetilde{\mathcal{D}}}^{-1}(U)$. We may assume, after passing to a subsequence, that the sequence $\{x_i\}$ converges. Let x denote the limit. By construction, $x \in \bigcap B_i$ and so $\pi_{\widetilde{\mathcal{D}}}(x) \in U$. However, the sequence $\{x_i\}$ is contained in the compact set $X \smallsetminus \pi_{\widetilde{\mathcal{D}}}^{-1}(U)$ and therefore so is the limit x. Then $\pi_{\widetilde{\mathcal{D}}}(x) \notin U$ and we have reached the desired contradiction.

Define the homeomorphism h as radially contracting each thickened disc of $\widetilde{D}^{s+1}$ into a small ball in its interior of radius less than ε. This uses the fact that the decomposition elements are properly nested. By construction, the homeomorphism h satisfies the two conditions listed above, the first by the choice of s and the second since each element of $\widetilde{\mathcal{D}}$ lies within $\widetilde{D}^{s+1}$, and $\operatorname{diam}_X h(\widetilde{D}_i^{s+1}) < \varepsilon$ for each thickened disc $\widetilde{D}_i^{s+1}$ of $\widetilde{D}^{s+1}$.

By the Bing shrinking criterion, we obtain a homeomorphism from $\big((D^2 \times D^2) \cup_{\mathrm{Id}} (S^3 \times [0,1])\big)/\widetilde{\mathcal{D}}$ to $D^2 \times D^2 \cup_{\mathrm{Id}} S^3 \times [0,1]$. Since $\widetilde{\mathcal{D}} = \mathcal{D}$, it follows that

$$\big((D^2 \times D^2) \cup_{\mathrm{Id}} (S^3 \times [0,1])\big)/\mathcal{D} \cong D^2 \times D^2 \cup_{\mathrm{Id}} S^3 \times [0,1].$$

The latter space is homeomorphic to $D^2 \times D^2$. $\square$

CHAPTER 26

Key facts about skyscrapers and decomposition space theory

Mark Powell and Arunima Ray

We record the key properties possessed by gropes, finite and infinite towers, and skyscrapers from Part II, along with the input from decomposition space theory from Part I, that we will need to complete the proof of the disc embedding theorem.

Here are the main facts we will need in the remaining step of the proof.

(1) The shrinking of mixed ramified Bing-Whitehead decompositions of $D^2 \times S^1$, from Chapter 8. The result is due to Ancel-Starbird [**AS89**] (see also Wright [**Wri89**] and Kasprowski-Powell [**KP14**]).

(2) The starlike null theorem, which states that null, recursively starlike equivalent decompositions of $D^2 \times D^2$ shrink, from Chapter 9. Closely related results are given in [**Bea67**, **DS83**, **FQ90**, **Dav07**, **MOR19**].

(3) The ball to ball theorem from Chapter 10. This result is due to Freedman [**Fre82a**], and was also proven in [**FQ90**, **Anc84**, **Sie82**].

(4) The skyscraper embedding theorem, which shows that skyscrapers may be embedded within any level of a given skyscraper, from Chapter 18. This result is due to Freedman-Quinn [**FQ90**, Chapters 1-3].

(5) The collar adding lemma from Chapter 25, which comes from [**FQ90**, p. 80].

As indicated by the list above, we will invoke all of the tools we have learnt so far. Next is a complete list of what we will need in the remaining two chapters, including more details on the results mentioned above.

26.1. Ingredients from Part I

THEOREM 26.1 (Shrinking of mixed ramified Bing-Whitehead decompositions (Theorem 8.1)). *A mixed ramified Bing-Whitehead decomposition of $D^2 \times S^1$ defined by a sequence of nested solid tori, embedded via either ramified Bing doubling or ramified Whitehead doubling of the cores $\{0\} \times S^1$, with b_i Bing embeddings between the $(i-1)$th and ith occurrence of a Whitehead double embedding, is shrinkable if and only if the series $\sum_i \frac{b_i}{2^i}$ diverges.*

SCHOLIUM 26.2 (Scholium 8.13). *Let $\mathcal{D}_{BW}$ be a mixed ramified Bing-Whitehead decomposition of $D^2 \times S^1$ defined by a sequence of nested solid tori as in Theorem 26.1 that shrinks, i.e. such that $\sum_j \frac{b_j}{2^j}$ diverges. Then there is a sequence of homeomorphisms $h_i \colon D^2 \times S^1 \to D^2 \times S^1$, for $i \geq 1$, with each h_i isotopic to the identity via an isotopy supported in the ith stage of the defining sequence for $\mathcal{D}_{BW}$, such that*

$$\lim_{m\to\infty} h_m \circ \cdots \circ h_1 \colon D^2 \times S^1 \to D^2 \times S^1$$

exists and has inverse sets coinciding with the quotient map $\pi\colon D^2 \times S^1 \to D^2 \times S^1/\mathcal{D}_{BW}$. *In particular, the defining sequence of the mixed ramified Bing-Whitehead decomposition can be repositioned so that each element of the new decomposition is a single point.*

Recall from Chapter 9 that a *red blood cell* in a 4-manifold M is constructed as a union $S^1 \times D^3 \cup D^2$, where $S^1 \times D^3 \subseteq M$, and D^2 is a flat embedded disc $D^2 \subseteq \overline{M \setminus (S^1 \times D^3)}$, such that $\partial D^2 = S^1 \times \{1\} \subseteq S^1 \times \partial D^3 \cong S^1 \times S^2$, and such that the flat neighbourhood $D^2 \times D^2$ restricts to a flat neighbourhood in $S^1 \times S^2 = S^1 \times \partial D^3$ for $S^1 \times \{1\}$.

Lemma 26.3 (Red blood cells are recursively starlike-equivalent (Lemma 9.11)). *Let* $E := (S^1 \times D^3) \cup D^2$ *be a red blood cell in a 4-manifold* M. *If* $S^1 \times D^3$ *has a product neighbourhood* $S^1 \times S^2 \times [0,1]$ *of its boundary embedded in* M, *then* E *is recursively starlike-equivalent.*

Theorem 26.4 (Starlike null theorem (Theorem 9.18)). *Let* $\mathcal{D}$ *be a null decomposition of* $D^2 \times D^2$ *with recursively starlike-equivalent decomposition elements, each of fixed filtration length* $K \geq 0$. *Suppose the decomposition elements are disjoint from the boundary of* $D^2 \times D^2$. *Also suppose that* $S^1 \times D^2 \subseteq \partial(D^2 \times D^2)$ *has a closed neighbourhood* C *disjoint from the decomposition elements. Then the quotient map* $\pi\colon D^2 \times D^2 \to (D^2 \times D^2)/\mathcal{D}$ *is approximable by homeomorphisms, agreeing with* π *on* $C \cup \partial(D^2 \times D^2)$. *In particular, there is a homeomorphism of pairs*

$$h\colon (D^2 \times D^2, S^1 \times D^2) \xrightarrow{\cong} ((D^2 \times D^2)/\mathcal{D}, S^1 \times D^2)$$

restricting to the quotient map π *on* $C \cup \partial(D^2 \times D^2)$.

Theorem 26.5 (Ball to ball theorem (Theorem 10.1)). *Let* $f\colon D^4 \to D^4$ *be a continuous map restricting to a homeomorphism* $f|_{S^3}\colon S^3 \to S^3$, *and let* E *be a closed subset of* D^4 *containing* S^3. *Suppose that the following holds.*

(a) The collection of inverse sets of f *is null.*
(b) The singular image of f *is nowhere dense.*
(c) The map f *restricts to a homeomorphism* $f|_{f^{-1}(E)}\colon f^{-1}(E) \to E$.

Then f *can be approximated by homeomorphisms that agree with* f *on* $f^{-1}(E)$.

26.2. Ingredients from Part II

Remark 26.6 (Chapter 14).

- For every open skyscraper $\mathcal{S}$ and for every $j \geq 1$, the truncated vertical boundary $\partial_+\mathcal{S}^{\leq j}$ is diffeomorphic to the complement of a neighbourhood of a mixed ramified Bing-Whitehead link in $D^2 \times S^1$ with Whitehead doubles in its last stage. Moreover, for each two consecutive truncations, the neighbourhoods are nested.
- There exist Kirby diagrams of (finite) towers corresponding to mixed ramified Bing-Whitehead links and of infinite towers corresponding to mixed ramified Bing-Whitehead decompositions.
- The vertical boundary of a skyscraper is homeomorphic to a solid torus and the total boundary is homeomorphic to S^3.

Chapter 13 explained how to describe towers and skyscrapers as subsets of $D^2 \times D^2$. This was summarised in Theorem 13.2. We give the essential facts here.

Proposition 26.7 (Theorem 13.2, Section 13.6). *Every finite truncation of a skyscraper, as well as every open skyscraper, embeds as a smooth submanifold with corners into* $D^2 \times D^2$, *such that the attaching region coincides with* $S^1 \times D^2$.

Moreover, the image of the embedding $\Psi_k \colon \mathcal{S}^{\leq k} \to D^2 \times D^2$ of a finite truncation can be described by a Kirby diagram given by a dotted m_k-component unlink in $S^3 = \partial(D^2 \times D^2)$, that is contained in $D^2 \times S^1$, obtained by mixed ramified Bing and Whitehead doubling the core curve $\{0\} \times S^1$, as described in Section 13.6. Thus $\Psi_k(\mathcal{S}^{\leq k})$ coincides with the complement in $D^2 \times D^2$ of an embedding

$$\Phi_k \colon \sqcup_{i=1}^{m_k} (D^2 \times \mathring{D}^2, S^1 \times D^2) \hookrightarrow (D^2 \times D^2, D^2 \times S^1)$$

of standard thickened slice discs for the m_k-component unlink. For the truncation $\mathcal{S}^{\leq k+1}$, the image of the embedding Φ_{k+1} lies inside the image of Φ_k, that is $\Psi_k(\mathcal{S}^{\leq k}) \subseteq \Psi_{k+1}(\mathcal{S}^{\leq k+1})$. The open skyscraper $\mathcal{S}$ has image the union

$$\bigcup_{k=1}^{\infty} \Psi_k(\mathcal{S}^{\leq k}) = (D^2 \times D^2) \smallsetminus \bigcap_{k=1}^{\infty} \operatorname{Im} \Phi_k \subseteq D^2 \times D^2.$$

PROPOSITION 26.8 (Proposition 12.20, Remark 12.22). *Every skyscraper is metrisable.*

REMARK 26.9 (Remark 12.19). For every open skyscraper $\mathcal{S}$, there is a canonical bijective correspondence between the ends of $\partial_+\mathcal{S}$ and the ends of $\mathcal{S}$. The endpoint compactification of $\partial_+\mathcal{S}$ is a subset of the endpoint compactification $\widehat{\mathcal{S}}$ of $\mathcal{S}$.

THEOREM 26.10 (Skyscraper embedding theorem (Theorem 18.10)). *Let $\widehat{\mathcal{S}}$ be a skyscraper. Let $\mathcal{T}_2$ be a connected component of some level of $\widehat{\mathcal{S}}$. Then there is an embedding of a skyscraper $\widehat{\mathcal{S}}' \subseteq \mathcal{T}_2$ such that the attaching region $\partial_-\widehat{\mathcal{S}}'$ and the first surface stage of $\widehat{\mathcal{S}}'$ agree with those of $\mathcal{T}_2$, and the connected components of the levels of $\widehat{\mathcal{S}}'$ with indices greater than or equal to 2 lie in arbitrarily small, mutually disjoint balls.*

LEMMA 26.11 (Collar adding lemma (Lemma 25.1)). *Let $\widehat{\mathcal{S}}$ be a skyscraper. Then the space*

$$\widehat{\mathcal{S}} \cup_{\mathrm{Id}} \big(\partial\widehat{\mathcal{S}} \times [0,1]\big),$$

glued along $\partial\widehat{\mathcal{S}} \times \{0\}$, is homeomorphic to $D^2 \times D^2$.

CHAPTER 27

Skyscrapers are standard: an overview

Stefan Behrens

We are ready for the final step in the proof of the disc embedding theorem. In Part II, we showed how to start with the hypotheses of the disc embedding theorem and build a mutually disjoint collection of skyscrapers. In the next, final chapter, we will show that skyscrapers, strange as they seem, are in fact homeomorphic to standard 2-handles, relative to the attaching region. As explained in Section 18.4, this will complete the proof of the disc embedding theorem. More precisely, we will prove the following theorem.

THEOREM 27.1. *For every skyscraper $\widehat{\mathcal{S}}$ there is a homeomorphism of pairs*

$$(F, \partial F) \colon \big(\widehat{\mathcal{S}}, \partial_- \widehat{\mathcal{S}}\big) \cong \big(D^2 \times D^2, S^1 \times D^2\big)$$

that is a diffeomorphism on a collar of $\partial_- \widehat{\mathcal{S}}$, such that if $\Phi \colon S^1 \times D^2 \to \partial_- \widehat{\mathcal{S}}$ is the attaching region, then $\partial F \circ \Phi = \mathrm{Id}_{S^1 \times D^2}$.

As we will point out in the next chapter, the proof we give will apply with only minor modification for the more general notion of skyscrapers introduced in Chapter 24.

For the remainder of this book, we reserve the symbol $\widehat{\mathcal{S}}$ for a given fixed skyscraper and the symbol H to denote the standard 2-handle $D^2 \times D^2$, which will be the target of the claimed homeomorphism from $\widehat{\mathcal{S}}$. The proof of Theorem 27.1 will take up the entire next chapter. Since the argument is rather complicated, in terms of mathematics as well as notational bookkeeping, in this chapter we give an overview of the upcoming proof.

27.1. An outline of the strategy

We will find subsets $\mathfrak{D} \subseteq \widehat{\mathcal{S}}$ and $D \subseteq H$ together with a homeomorphism $\varphi \colon \mathfrak{D} \to D$. This abstract object is called *the design*, while $\mathfrak{D}$ and D are called *the design in the skyscraper* and *the design in the standard handle* respectively. Note that there is not a unique design, so the definite article *the* only applies once these subsets have been found and fixed. Call each closure of a connected component of the complement of $\mathfrak{D}$ in $\widehat{\mathcal{S}}$ a *gap*, and each closure of a connected component of the complement of D in H a *hole*. Ideally, we would be able to quotient out by the decompositions given by the gaps and the holes respectively, find homeomorphic spaces, and use decomposition space theory to conclude that the original spaces were homeomorphic to begin with. However, we will need a further step. We will have to expand each gap and each hole a bit so that the complements of the union of each family are still homeomorphic. This process expands each gap to a *gap*$^+$, and each hole to a *hole*$^+$. Write $\mathcal{G}^+$ and $\mathcal{H}^+$ for the decompositions of $\widehat{\mathcal{S}}$ and H given by the holes$^+$ and the gaps$^+$ respectively. We will show the following facts.

(1) The quotients $\widehat{\mathcal{S}}/\mathcal{G}^+$ and $H/\mathcal{H}^+$ are homeomorphic.

(2) The decomposition $\mathcal{H}^+$ is shrinkable, that is, the quotient map $\alpha\colon H \to H/\mathcal{H}^+$ is approximable by homeomorphisms. This is called the α *shrink*.
(3) The decomposition $\mathcal{G}^+$ is shrinkable, that is, the quotient map $\beta\colon \widehat{\mathcal{S}} \to \widehat{\mathcal{S}}/\mathcal{G}^+$ is approximable by homeomorphisms. This is the β *shrink*.

From these statements, we shall conclude that there is a sequence of homeomorphisms

$$\begin{tikzcd} \widehat{\mathcal{S}} \arrow[d, "\cong"] & H \arrow[d, "\cong"] \\ \widehat{\mathcal{S}}/\mathcal{G}^+ \arrow[r, "\cong"] & H/\mathcal{H}^+ \end{tikzcd}$$

relative to the attaching region. This will complete the proof.

27.2. The strategy in more detail

First we discuss the construction of $\mathfrak{D} \subseteq \widehat{\mathcal{S}}$. Skyscrapers have three key properties:

(i) their finite truncations can be embedded in the standard handle (Proposition 26.7);
(ii) every skyscraper contains many other embedded skyscrapers whose upper levels are suitably small and separated from one another (Theorem 26.10); and
(iii) the vertical boundaries of skyscrapers are very well understood (Remark 26.6).

We will use the embedding property (ii) to iteratively find uncountably many skyscrapers embedded within $\widehat{\mathcal{S}}$, and $\mathfrak{D}$ will consist of collars of the vertical boundaries of these embedded skyscrapers. The collection of embedded skyscrapers will be indexed by finite and infinite sequences of 0s and 1s. Let I be such a sequence; we denote the corresponding skyscraper by $\widehat{\mathcal{S}}_I$. For the empty word $I = \varnothing$, the skyscraper $\widehat{\mathcal{S}}_\varnothing$ will be defined to be $\widehat{\mathcal{S}}$ itself. Recall that $\mathcal{S}_I$ denotes the open skyscraper corresponding to $\widehat{\mathcal{S}}_I$. For each finite I, we define a smooth boundary collar

$$c_I\colon \partial_+\mathcal{S}_I \times [0,1] \hookrightarrow \mathcal{S}_I,$$

with $\partial_+\mathcal{S}_I \times \{0\}$ mapped to $\partial_+\mathcal{S}_I$ and such that the collars for the different (open) skyscrapers are compatible with one another. The design $\mathfrak{D} \subseteq \widehat{\mathcal{S}}$ will be the union of pieces from the various embedded skyscrapers. Recall that $\mathcal{S}_I^\ell$ denotes the ℓth level of the open skyscraper $\mathcal{S}_I$. For finite I of length $k \neq 0$, let $\mathfrak{D}_I = c_I\big(\partial_+\mathcal{S}_I^{k+1} \times [0,1]\big)$, while $\mathfrak{D}_\varnothing$ is defined to be the union of $c_\varnothing\big(\partial_+\mathcal{S}_\varnothing^1 \times [0,1]\big)$ and a collar of $\partial_-\widehat{\mathcal{S}}_\varnothing = \partial_-\widehat{\mathcal{S}}$. For each infinite I, let $\mathfrak{D}_I$ be the set of endpoints of $\widehat{\mathcal{S}}_I$. The union of the above pieces is the design $\mathfrak{D}$ in the skyscraper, that is

$$\mathfrak{D} := \bigcup_I \mathfrak{D}_I \subseteq \widehat{\mathcal{S}}.$$

By the compatibility of collars in the construction, it will be clear that for each infinite I the full vertical boundary of $\widehat{\mathcal{S}}_I$ is contained in $\mathfrak{D}$, along with a collar for $\partial_+\mathcal{S}_I$, which gets progressively thinner as we climb higher in $\widehat{\mathcal{S}}_I$. Recall that the vertical boundary of any skyscraper is homeomorphic to a solid torus. See Figure 27.1.

Next we need to find the design in the standard 2-handle H. Let $\widehat{\mathcal{S}}_I$ be an embedded skyscraper in $\widehat{\mathcal{S}}$ corresponding to an infinite sequence I. By Remark 26.6, the vertical boundary of any finite truncation $\widehat{\mathcal{S}}_I^{\leq k}$ is the exterior of some mixed ramified Bing-Whitehead link L_I^k in a solid torus $D^2 \times S^1$. Note that the standard

2-handle $H = D^2 \times D^2$ decomposes as $\big(D^2 \times S^1 \times [\frac{2}{3}, 1]\big) \cup \big(D^2 \times D^2_{\frac{2}{3}}\big)$. That is, away from a central $D^2 \times D^2$, H decomposes as an interval's worth of solid tori. We will embed each link L_I^k in solid tori $D^2 \times S^1 \times \{u_I\}$, for u_I ranging over certain subintervals $J_I \subseteq [\frac{2}{3}, 1]$. The intervals J_I correspond to removed middle-thirds in the Cantor set construction on $[\frac{2}{3}, 1]$, so one has embeddings $L_I^k \times J_I \subseteq D^2 \times S^1 \times J_I$. Fitting these embeddings together carefully will lead to finding the design D in H. The identification $\varphi \colon \mathfrak{D} \to D$ will be made painfully explicit later on. For now we observe that since the full vertical boundary of each reimbedded skyscraper corresponding to an infinite sequence is contained within $\mathfrak{D}$, the design D in H contains uncountably many solid tori at radii corresponding to the Cantor set in a precise way. A schematic diagram of the design in $\widehat{\mathcal{S}}$ and the design in H is given in Figure 27.1.

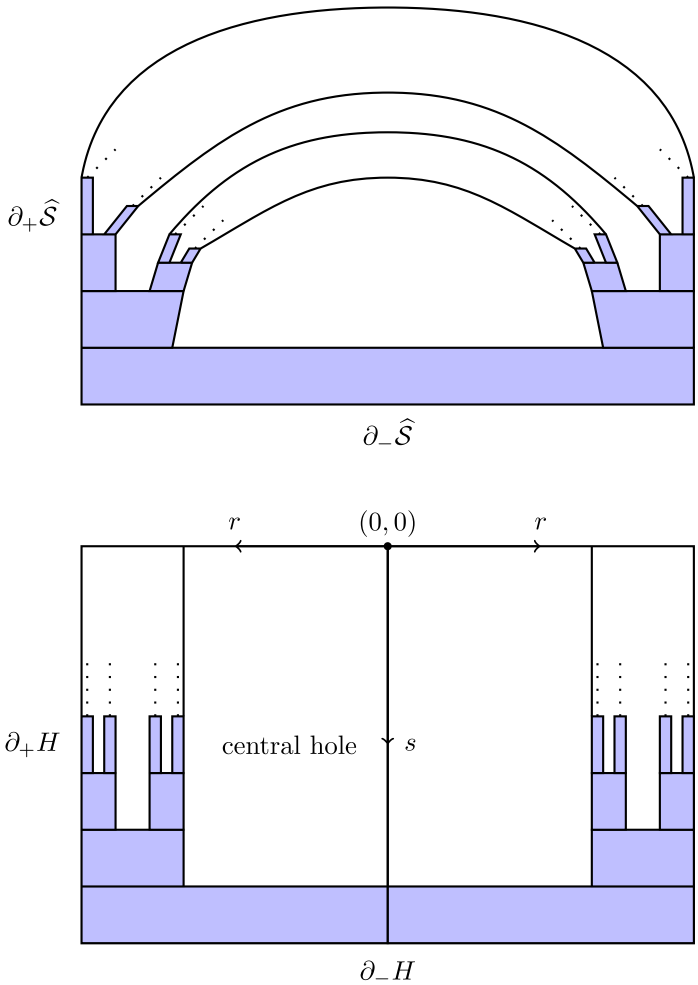


Figure 27.1. The design inside the skyscraper $\widehat{\mathcal{S}}$ (top) and the design inside the standard handle H (bottom). The bottom picture uses coordinates $(s, \theta; r, \varphi)$, namely polar coordinates in the two factors of $D^2 \times D^2$. The θ coordinate is suppressed and only $\varphi = 0, \pi$ are shown.

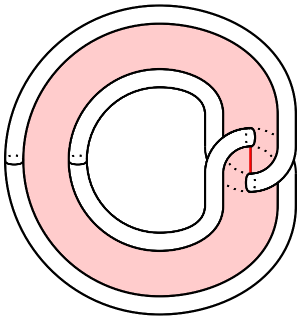


Figure 27.2. An immersed disc bounded by a component of the link L_I^k.

As mentioned earlier, each closure of a connected component of the complement of $\mathfrak{D}$ in $\widehat{\mathcal{S}}$ is called a gap, and each closure of a connected component of the complement of D in H is called a hole. The homeomorphism types of the gaps in $\widehat{\mathcal{S}}$ are highly mysterious. However, the situation in H is completely understood, given our explicit and careful embedding of the design there. Since the design pieces will correspond to the exteriors of links, the components of the complement of the design are neighbourhoods of components of such links, crossed with an interval. That is, other than a central hole diffeomorphic to $D^2 \times D^2$, each hole is diffeomorphic to $S^1 \times D^3$. This immediately shows why we cannot shrink the decompositions corresponding to holes and gaps. Since the holes are not cellular, they do not shrink (Proposition 4.21). We will address this problem by expanding the holes and gaps to holes$^+$ and gaps$^+$ respectively, as follows. Each component of a mixed ramified Bing-Whitehead link L_I^k that appears in the design D in H bounds an immersed disc in a solid torus $D^2 \times S^1 \times \{t_I\}$, where t_I is the midpoint of the interval J_I, as shown in Figure 27.2. Moreover, recall that there are uncountably many complete solid tori in the design D in H, corresponding to the full vertical boundaries of the embedded skyscrapers in $\widehat{\mathcal{S}}$ associated with infinite sequences of 0s and 1s. The immersed discs bounded by the holes lie in 3-dimensional slices of H. Using the extra dimension we are able to slightly perturb them into embedded discs in H, which provide *red blood cell discs*. That is, add one such disc to each hole to produce the collection of holes$^+$. With some extra care, we can ensure that the red blood cell discs for different holes are mutually disjoint and we perturb them into the full vertical boundaries, so that the red blood cell discs live inside the design D in H. The latter fact allows us to pull them over to $\widehat{\mathcal{S}}$ via the homeomorphism $\varphi\colon \mathfrak{D} \to D$, where we use them to form the gaps$^+$. We note that each hole$^+$ is contractible and even turns out to be cellular, which improves our chances of being able to shrink the decomposition.

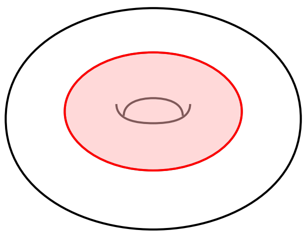


Figure 27.3. A schematic picture of a hole$^+$ showing a red blood cell disc (red). The fourth dimension of the $S^1 \times D^3$ piece is suppressed.

In the rest of the proof, we show that the quotients $\widehat{\mathcal{S}}/\mathcal{G}^+$ and $H/\mathcal{H}^+$ are homeomorphic, that the decomposition $\mathcal{H}^+$ of H is shrinkable (the α shrink), and that the decomposition $\mathcal{G}^+$ of $\widehat{\mathcal{S}}$ is shrinkable (the β shrink).

For the first statement, we will appeal to the fact that the gaps$^+$ and the holes$^+$ are uniquely determined by their intersection with $\mathfrak{D}$ and D respectively. This will show that

$$\widehat{\mathcal{S}}/\mathcal{G}^+ \cong \mathfrak{D}/\mathcal{G}^+ \text{ and } D/\mathcal{H}^+ \cong H/\mathcal{H}^+,$$

where, by a mild abuse of notation, the symbol $\mathfrak{D}/\mathcal{G}^+$ refers to the quotient $\mathfrak{D}/(\mathfrak{D}\cap\mathcal{G}^+)$, and similarly $D/\mathcal{H}^+$ denotes the quotient $D/(D\cap\mathcal{H}^+)$. Moreover, by construction we have a homeomorphism $\mathfrak{D}/\mathcal{G}^+ \cong D/\mathcal{H}^+$ induced by $\varphi\colon \mathfrak{D} \xrightarrow{\cong} D$.

For the α shrink, we will exploit our complete understanding of $\mathcal{H}^+$ and apply Theorem 26.4. This theorem applies to null decompositions whose elements are recursively starlike-equivalent of fixed filtration length. Each hole$^+$ is recursively starlike-equivalent of filtration length 1 by construction (Lemma 26.3) and it is not hard to arrange the collection to be null.

For the β shrink we will use the ball to ball theorem (Theorem 26.5). This seems counterintuitive since the theorem requires the domain of the function under consideration to be a 4-ball, which is what we are trying to prove in the first place. We arrange for this by adding a collar to both the domain and codomain, namely to the skyscraper $\widehat{\mathcal{S}}$ and to the quotient $\widehat{\mathcal{S}}/\mathcal{G}^+$, and extending the function by the identity. By the collar adding lemma (Lemma 26.11), we see that the new domain is a 4-ball. The codomain is a 4-ball since $\widehat{\mathcal{S}}/\mathcal{G}^+ \cong H/\mathcal{H}^+$, and we have already shown that the latter space $H/\mathcal{H}^+$ is a 4-ball by the α shrink. In order to apply the ball to ball theorem, we have to check that $\mathcal{G}^+$ is null and the image of the gaps$^+$ is nowhere dense. The nullity of $\mathcal{G}^+$ is shown directly, based on how the design embeds in $\widehat{\mathcal{S}}$. The argument uses that the uncountably many skyscrapers embedded within $\widehat{\mathcal{S}}$ have their upper levels suitably squeezed and separated from one another. To see that the image of $\mathcal{G}^+$ in the quotient $\widehat{\mathcal{S}}/\mathcal{G}^+$ is nowhere dense, it will suffice to show that the image of $\mathcal{H}^+$ in $D/\mathcal{H}^+$ is nowhere dense, and this turns out to be relatively easy to verify. Then the ball to ball theorem shows that the map between balls can be approximated by a homeomorphism that restricts to the given map on the collars, implying that the spaces without the collars, namely our original skyscraper and a 4-ball, are homeomorphic relative to the attaching region.

27.3. Some things to keep in mind

In the next chapter, we will fill in all the details for the steps outlined above and give the proof of Theorem 27.1 in its full glory. We finish this chapter by reminding the reader of some concrete objectives one should keep in mind in our construction of the design in the skyscraper, the design in the standard handle, the holes$^+$, and the gaps$^+$.

(1) We wish to embed uncountably many skyscrapers within the initially given skyscraper $\widehat{\mathcal{S}}$, indexed by finite and infinite sequences of 0s and 1s, and define the design in $\widehat{\mathcal{S}}$ to consist of tapering collars (including endpoints) of the vertical boundaries of skyscrapers corresponding to infinite binary words. The full vertical boundary, a solid torus, of each such skyscraper must be in the design.

(2) We wish to find an embedding of the design in the standard handle H. A helpful tool for this is our understanding of Kirby diagrams for truncations of skyscrapers, consisting of mixed ramified Bing-Whitehead links. We will compatibly embed these links within solid tori slices of H.

(3) When expanding holes to holes$^+$, red blood cell discs need to be found disjointly within the design, utilising the full solid tori corresponding to the vertical boundaries of embedded skyscrapers, which are contained within the design.
(4) In order to identify the quotients $\widehat{\mathcal{S}}/\mathcal{G}^+$ and $H/\mathcal{H}^+$ we must ensure that the gaps$^+$ and the holes$^+$ are uniquely determined by their intersections with $\mathfrak{D}$ and D, respectively.
(5) For the α shrink we need the collection of holes$^+$ to be null. This should inform how we embed skyscrapers within $\widehat{\mathcal{S}}$ and how we find the red blood cell discs.
(6) For the β shrink using the ball to ball theorem, we need the collection of gaps$^+$ to be null and the image of the gaps$^+$ in the quotient $\widehat{\mathcal{S}}/\mathcal{G}^+$ to be nowhere dense. This should also inform how we embed skyscrapers within $\widehat{\mathcal{S}}$.

CHAPTER 28

Skyscrapers are standard: the details

Stefan Behrens, Daniel Kasprowski, Mark Powell, and Arunima Ray

We give a detailed proof that every skyscraper $\widehat{\mathcal{S}}$ is homeomorphic to $D^2 \times D^2$ relative to its attaching region (Theorem 27.1). We follow the structure of the proofs of [**Fre82a**, **FQ90**], adding more details. For the entire chapter, fix a skyscraper $\widehat{\mathcal{S}}$ and denote the standard handle $D^2 \times D^2$ by H.

28.1. Binary words and the Cantor set

We begin by introducing the main bookkeeping device for the construction of the design outlined in the previous chapter. Let $\overline{\mathbb{N}}$ denote the set $\mathbb{N} \cup \{0, \infty\}$. A *binary word* is a finite (possibly empty) or infinite sequence $I = i_1 i_2 \dots i_k \dots$ of 0s and 1s, that is an element of $\{0,1\}^k$ where $k \in \overline{\mathbb{N}}$. The symbol $|I|$ denotes the length of I. The *empty word* is denoted by $\varnothing$ and has length 0. For finite k and possibly infinite ℓ we have the juxtaposition operation, denoted by $\cdot$ or often simply by concatenation, given by

$$\begin{aligned} \{0,1\}^k \times \{0,1\}^\ell \longrightarrow & \{0,1\}^{k+\ell} \\ (i_1 \dots i_k, j_1 \dots) \longmapsto & i_1 \dots i_k \cdot j_1 \dots \\ = & i_1 \dots i_k j_1 \dots \end{aligned}$$

If J is a binary word and I is a finite binary word, we say that J *starts with* I, denoted by $I \preceq J$, if $J = I \cdot I'$ for some binary word I'. This defines a partial order on the set of binary words. We say that an infinite binary word is *essentially finite* if it has an infinite tail of either 0s or 1s, that is if it is of the form $I \cdot 000\dots$ or $I \cdot 111\dots$ for some finite binary word I. Note that the set of finite binary words is countably infinite since it is the (countable) union of the finite sets of binary words of a given finite length. On the other hand, the set of infinite binary words is uncountable and has the same cardinality as the power set of $\mathbb{N}$.

Binary words have a well known correspondence to the *ternary Cantor set* $\mathfrak{C}_3 \subseteq [0,1]$, which will play a prominent rôle in our constructions. Recall that $\mathfrak{C}_3$ is obtained by removing the open 'middle third' subinterval first from $[0,1]$, then from each of $[0, \frac{1}{3}]$ and $[\frac{2}{3}, 1]$, and so on from the remaining intervals *ad infinitum* (see Figure 28.1). The dyadic branching in each step of this construction is reminiscent of the process of extending a binary word of length $k < \infty$ to one of length $k+1$ by adding either a 0 or a 1. At each stage, we have a choice of adding a 0 (turning left) or adding a 1 (turning right). Thus finite binary words of fixed length k correspond to the subintervals of length $\frac{1}{3^k}$ that survive in the kth step of constructing $\mathfrak{C}_3$, and we may index these subintervals as

$$C_\varnothing = [0,1],$$

$$C_0 = \left[0, \frac{1}{3}\right], \quad C_1 = \left[\frac{2}{3}, 1\right],$$

$$C_{00} = \left[0, \frac{1}{9}\right], \quad C_{01} = \left[\frac{2}{9}, \frac{1}{3}\right], \quad C_{10} = \left[\frac{2}{3}, \frac{7}{9}\right], \quad C_{11} = \left[\frac{8}{9}, 1\right],$$

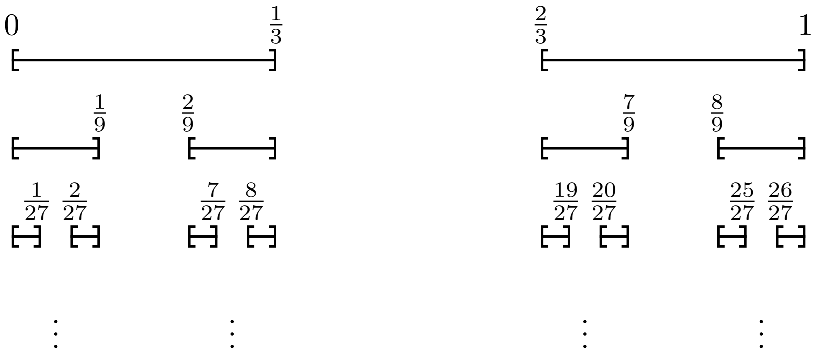


Figure 28.1. Constructing the ternary Cantor set.

and so on (see Figure 28.1). Note that these subintervals get smaller and smaller with each iteration. The points that survive in the limit of this process, namely the points in $\mathfrak{C}_3$, correspond to infinite binary words. This relationship is formalised by the following function. For $k \in \overline{\mathbb{N}}$, define

$$\begin{aligned} \mathfrak{c}\colon \{0,1\}^k &\longrightarrow \mathfrak{C}_3 \\ i_1 i_2 \cdots &\longmapsto \sum_{j=1}^{k} \frac{2\,i_j}{3^j}. \end{aligned}$$

For a binary word I, let $\mathfrak{c}_I$ denote the quantity

$$\mathfrak{c}_I := \mathfrak{c}(I) = \sum_{j=1}^{|I|} \frac{2\,i_j}{3^j}.$$

Note that $\mathfrak{c}_\varnothing = 0$. One should check that $\mathfrak{c}$ maps into $\mathfrak{C}_3$ and is a bijection when restricted to infinite binary words. Indeed, for a finite binary word I, $\mathfrak{c}_I$ is the left boundary point of the interval C_I defined above, which is thus described as

$$C_I = \left[\mathfrak{c}_I,\ \mathfrak{c}_I + \tfrac{1}{3^{|I|}}\right] \subseteq [0,1].$$

It is routine to check that

$$\mathfrak{c}_{I\cdot 00\ldots} = \mathfrak{c}_I \quad \text{and} \quad \mathfrak{c}_{I\cdot 11\ldots} = \mathfrak{c}_I + \tfrac{1}{3^{|I|}}.$$

Thus $\mathfrak{c}$ maps essentially finite binary words to the boundary points of the intervals $\{C_I\}$ and we have the following affine parametrisation for any finite binary word I.

$$\begin{aligned} a_I\colon [0,1] &\longrightarrow C_I \\ t &\longmapsto \mathfrak{c}_I + \frac{1}{3^{|I|}} t \\ &= \sum_{j=1}^{|I|} \frac{2\,i_j}{3^j} + \frac{1}{3^{|I|}} t. \end{aligned}$$

The next two lemmas will be used in the proof of Lemma 28.20.

Lemma 28.1. *Let I and J be finite binary words with $I \neq J$. Then either $\mathfrak{c}_{I\cdot 1} < \mathfrak{c}_{J\cdot 0111\ldots}$ or $\mathfrak{c}_{J\cdot 1} < \mathfrak{c}_{I\cdot 0111\ldots}$.*

PROOF. As previously observed, the points $\mathfrak{c}_{I\cdot 1}$ and $\mathfrak{c}_{J\cdot 1}$ are the left endpoints of the intervals $C_{I\cdot 1}$ and $C_{J\cdot 1}$ respectively, while the points $\mathfrak{c}_{I\cdot 0111\ldots}$ and $\mathfrak{c}_{J\cdot 0111\ldots}$ are the right endpoints of the intervals $C_{I\cdot 0}$ and $C_{J\cdot 0}$ respectively. Since the function $\mathfrak{c}$ restricted to infinite binary words is a bijection, if $\mathfrak{c}_{I\cdot 0111\ldots} = \mathfrak{c}_{J\cdot 0111\ldots}$, then $I = J$, which is not the case by hypothesis. So $\mathfrak{c}_{I\cdot 0111\ldots} \neq \mathfrak{c}_{J\cdot 0111\ldots}$. Similarly, if $\mathfrak{c}_{J\cdot 1} = \mathfrak{c}_{J\cdot 1000\ldots} = \mathfrak{c}_{I\cdot 1000\ldots} = \mathfrak{c}_{I\cdot 1}$, then $I = J$, which again is not the case. Thus the four points are distinct, since right endpoints and left endpoints never coincide.

Suppose that $\mathfrak{c}_{I\cdot 1} \geq \mathfrak{c}_{J\cdot 0111\ldots}$. We must show that $\mathfrak{c}_{J\cdot 1} < \mathfrak{c}_{I\cdot 0111\ldots}$. By the previous paragraph, $\mathfrak{c}_{I\cdot 1} > \mathfrak{c}_{J\cdot 0111\ldots}$, which implies that $\mathfrak{c}_{I\cdot 0111\ldots} > \mathfrak{c}_{J\cdot 0111\ldots}$, since there are no points of the Cantor set between $\mathfrak{c}_{I\cdot 0111\ldots}$ and $\mathfrak{c}_{I\cdot 1}$ and since $\mathfrak{c}_{I\cdot 0111\ldots} \neq \mathfrak{c}_{J\cdot 0111\ldots}$. Moreover, $\mathfrak{c}_{I\cdot 0111\ldots} - \mathfrak{c}_{J\cdot 0111\ldots} > \frac{1}{3^{|J|}}$, since there are no points of the Cantor set in the open interval of length $\frac{1}{3^{|J|}}$ to the right of the right endpoint $\mathfrak{c}_{J\cdot 0111\ldots}$, and equality is not possible since $\mathfrak{c}_{I\cdot 0111\ldots}$ is a right endpoint. We also know that $\mathfrak{c}_{J\cdot 1} - \mathfrak{c}_{J\cdot 0111\ldots} = \frac{1}{3^{|J|}}$. Thus

$$\mathfrak{c}_{J\cdot 1} = \mathfrak{c}_{J\cdot 0111\ldots} + \frac{1}{3^{|J|}} < \mathfrak{c}_{J\cdot 0111\ldots} + \mathfrak{c}_{I\cdot 0111\ldots} - \mathfrak{c}_{J\cdot 0111\ldots} = \mathfrak{c}_{I\cdot 0111\ldots}$$

as desired. $\square$

LEMMA 28.2. *Let I and J be finite binary words with $I \neq J$ and $\mathfrak{c}_{I\cdot 1} < \mathfrak{c}_{J\cdot 0111\ldots}$. Then there exists some $c \in \mathfrak{C}_3$ with $\mathfrak{c}_{I\cdot 1} < c < \mathfrak{c}_{J\cdot 0111\ldots}$.*

PROOF. Let k be an integer such that $\frac{1}{3^k} < \mathfrak{c}_{J\cdot 0111\ldots} - \mathfrak{c}_{I\cdot 1}$. Choose $m \geq 2$ so that $|J|+m > k$. Let K be the finite binary word $J\cdot 0111\ldots 1$ of length $|J|+m$. We claim that $c = \mathfrak{c}_K$ satisfies $\mathfrak{c}_{I\cdot 1} < c < \mathfrak{c}_{J\cdot 0111\ldots}$. Since $\mathfrak{c}_K < \mathfrak{c}_{J\cdot 0111\ldots}$ by construction, it suffices to show that $\mathfrak{c}_{J\cdot 0111\ldots} - \mathfrak{c}_K < \frac{1}{3^k}$, which in turn is less than $\mathfrak{c}_{J\cdot 0111\ldots} - \mathfrak{c}_{I\cdot 1}$ by the choice of k. And indeed

$$\mathfrak{c}_{J\cdot 0111\ldots} - \mathfrak{c}_K = \sum_{i=|J|+m+1}^{\infty} \frac{2}{3^i} = \frac{1}{3^{|J|+m}} < \frac{1}{3^k},$$

as needed. $\square$

Finally note that one can perform the ternary Cantor set construction on an interval of any length, rather than $[0,1]$. We will soon begin carefully building the design in the skyscraper and the standard handle. In doing so, we will often work with the intersection $\mathfrak{C}_3 \cap [\frac{2}{3}, 1]$. Due to the self-similarity inherent in the Cantor set, this intersection is also the result of performing the Cantor set construction on the interval $[\frac{2}{3}, 1]$.

28.2. The standard handle

Throughout this chapter, we will use a standard, explicit parametrisation of the 2-handle H using polar coordinates within each disc factor. To be precise, we write $H = \{(s,\theta;r,\varphi)\}$, where $0 \leq r,s \leq 1$ and $0 \leq \theta,\varphi \leq 2\pi$. Here $(s,\theta;r,\varphi)$ corresponds to $(se^{i\theta}, re^{i\varphi}) \in D^2 \times D^2 \subseteq \mathbb{C} \times \mathbb{C}$. We define

$$\partial_- H := S^1 \times D^2 = \{s = 1\} \quad \text{and} \quad \partial_+ H := D^2 \times S^1 = \{r = 1\}.$$

The parametrisation is shown in Figure 28.2, where the angle coordinates have been suppressed. The left hand edge of the figure, where $r = 1$, corresponds to the solid torus $\partial_+ H = D^2 \times S^1$. The bottom edge, where $s = 1$, corresponds to the solid torus $\partial_- H = S^1 \times D^2$. The top edge, where $s = 0$, corresponds to the cocore $\{0\} \times D^2$, while the right hand edge, where $r = 0$, corresponds to the core $D^2 \times \{0\}$. As expected, the core and cocore have a single point of intersection at the top right corner. Note that the vertical line segment at any radius $r \in (0,1]$ corresponds to

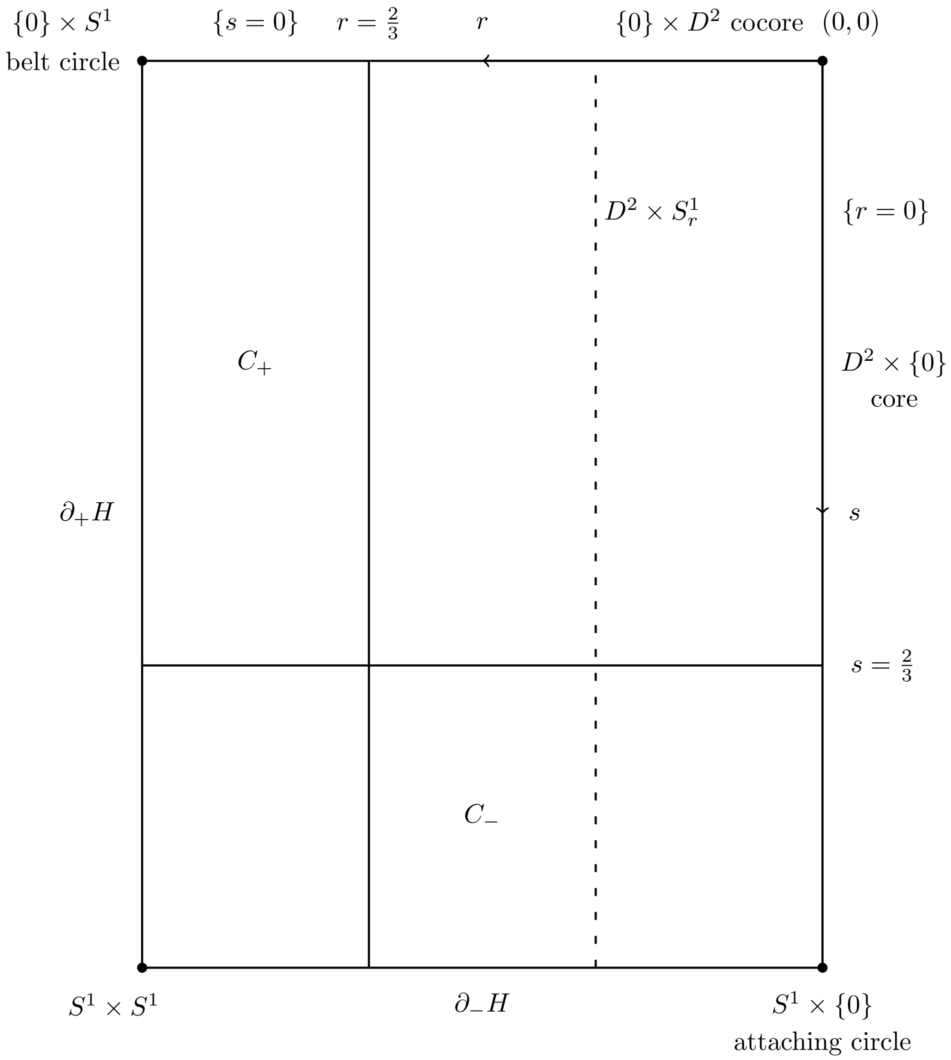


Figure 28.2. The parametrisation of $H := D^2 \times D^2 = \{(s, \theta; r, \varphi) \mid 0 \leq r, s \leq 1, 0 \leq \theta, \varphi \leq 2\pi\}$ with the angle coordinates suppressed. The right hand edge represents the core $D^2 \times \{0\}$, the top edge the cocore $\{0\} \times D^2$, the left hand edge $\partial_+ H = D^2 \times S^1$, and the bottom edge $\partial_- H = S^1 \times D^2$.

a solid torus $D^2 \times S^1_r$ at radius r. Our goal in this section is to embed the design $\mathfrak{D}$ in H.

We also define collars for the boundary of H, as follows (see Figure 28.2). Let

$$\begin{aligned} C_- \colon \partial_- H \times [0,1] = S^1 \times D^2 \times [0,1] &\to H = D^2 \times D^2 \\ ((1, \theta; r, \varphi), s) &\mapsto (1 - s/3, \theta; r, \varphi) \end{aligned}$$

and

$$\begin{aligned} C_+ \colon \partial_+ H \times [0,1] = D^2 \times S^1 \times [0,1] &\to H = D^2 \times D^2 \\ ((s, \theta; 1, \varphi), r) &\mapsto (s, \theta; 1 - r/3, \varphi). \end{aligned}$$

Remark 28.3. Here and throughout this chapter, the collars are such that the map restricted to $0 \in [0,1]$ is the identity. Note that C_- and C_+ are not collars in the traditional sense for manifold boundaries since, for example, $C_-((\partial_+H \cap \partial_-H) \times (0,1])$ is not contained in the interior of H but rather lies in ∂_+H; such collars are natural to consider in the study of manifolds with corners (see e.g. [**Wal16**, Section 2.6]). All collars we define from now on will be of this nature. In the language of [**Wal16**], the collars C_- and C_+ together form a *semicollar* for the boundary of H. Whenever a corner of a manifold M is 2-sided in the boundary of M, there exists a semicollar neighbourhood of the boundary [**Wal16**, p. 60].

The following lemma is immediate from the definition of the collars C_- and C_+.

LEMMA 28.4. *If there exist $p_- := (1, \theta_-; r_-, \varphi_-) \in \partial_-H$, $p_+ := (s_+, \theta_+; 1, \varphi_+) \in \partial_+H$, and $t_\pm \in [0,1]$ so that $C_-(p_-, t_-) = C_+(p_+, t_+)$, then*

$$\Big(1 - \frac{t_-}{3}, \theta_-; r_-, \varphi_-\Big) = \Big(s_+, \theta_+; 1 - \frac{t_+}{3}, \varphi_+\Big).$$

That is, $\theta_- = \theta_+$, $\varphi_- = \varphi_+$, $1 - \frac{t_-}{3} = s_+$, and $r_- = 1 - \frac{t_+}{3}$.

28.3. Embedding the design in a skyscraper

Our overarching goal is to find a suitable space embedded within both the skyscraper $\widehat{\mathcal{S}}$ and the standard 2-handle H. In this section, we find this subspace within $\widehat{\mathcal{S}}$. By Proposition 26.8 the skyscraper $\widehat{\mathcal{S}}$ is metrisable. Fix a metric on $\widehat{\mathcal{S}}$ that induces its topology.

The desired subspace of $\widehat{\mathcal{S}}$ will be produced from uncountably many other skyscrapers which we will find embedded within $\widehat{\mathcal{S}}$ in a compatible manner. These embedded skyscrapers will be indexed by (finite and infinite) binary words. Each skyscraper will have a collar of its vertical boundary, which will yield a *design piece* within $\widehat{\mathcal{S}}$. Consequently, design pieces will also be indexed by binary words. We will then embed these design pieces back into H, to give us a common subset of both $\widehat{\mathcal{S}}$ and H that we understand well. As explained in the previous chapter, our overall goal will then be to modify the common subset until we can shrink the decompositions determined by the closures of the connected components of the complement of the common subset.

Recall that the open skyscraper corresponding to some skyscraper $\widehat{\mathcal{S}}'$ is denoted by $\mathcal{S}'$. It will be useful in some steps of the argument to identify the open skyscraper $\mathcal{S}$ with a submanifold of H. By Proposition 26.7, the open skyscraper $\mathcal{S}$ is diffeomorphic to a submanifold of H such that the attaching regions of $\mathcal{S}$ and H coincide. This was done in Chapter 13 using compatible Kirby diagrams for the finite truncations of a skyscraper (see Remark 26.6). We stipulate the following conditions on this representation.

- (S1) For every finite truncation of $\mathcal{S}$, the dotted circles representing 1-handles in $S^3 = \partial H$ are disjoint from the image of C_-.
- (S2) The Kirby diagram for every finite truncation of $\mathcal{S}$ satisfies the repositioning criterion of Scholium 26.2. In other words, the elements of the corresponding mixed ramified Bing-Whitehead decomposition are singletons.
- (S3) The thickened discs removed from H, corresponding to the 1-handles of the finite truncations of $\mathcal{S}$, are products within the image of C_+, so that their intersections with every r-level set in C_+ is constant, i.e. independent of r.

To be more precise, for the rest of the chapter we fix a continuous map

$$\Gamma \colon \mathcal{S} \hookrightarrow H$$

that is a diffeomorphism onto its image, and such that if $\Phi\colon S^1 \times D^2 \to \partial_- \mathcal{S}$ is the attaching region, we have that

$$\Gamma \circ \Phi\colon S^1 \times D^2 \to \partial_- \widehat{\mathcal{S}} \to \partial_- H = S^1 \times D^2 \tag{28.1}$$

is the identity map. Conditions (S1) and (S3) will be used in the next section. Condition (S2) will be needed in Section 28.4.1 and for Lemma 28.14.

Next we begin to describe the design pieces in $\widehat{\mathcal{S}}$, one for each binary word.

28.3.1. The design piece for the empty word. Define the skyscraper

$$\widehat{\mathcal{S}}_\varnothing := \widehat{\mathcal{S}},$$

as well as the collars

$$\begin{aligned} c_-\colon \partial_- \mathcal{S}_\varnothing \times [0,1] &\to \mathcal{S}_\varnothing \\ (x,t) &\mapsto \Gamma^{-1}\big(C_-(\Gamma(x),t)\big) \end{aligned}$$

and

$$\begin{aligned} c_\varnothing\colon \partial_+ \mathcal{S}_\varnothing \times [0,1] &\to \mathcal{S}_\varnothing \\ (x,t) &\mapsto \Gamma^{-1}\big(C_+(\Gamma(x),t)\big). \end{aligned}$$

Here c_- is defined and gives a collar of $\partial_- \mathcal{S}_\varnothing$ in $\mathcal{S}_\varnothing$ by condition (S1) in the definition of Γ. Similarly, $c_\varnothing$ is a collar of $\partial_+ \mathcal{S}_\varnothing$ in $\mathcal{S}_\varnothing$ by condition (S3). In particular, since each of C_- and C_+ restricted to $0 \in [0,1]$ is the identity, the same applies to the collars c_- and $c_\varnothing$. As mentioned in Remark 28.3, the pair c_- and $c_\varnothing$ give a semicollar of $\partial \mathcal{S}_\varnothing$.

Moreover, condition (S3) in the definition of Γ implies that

$$c_\varnothing\big(\partial_+ \mathcal{S}^\ell_\varnothing \times [0,1]\big) \subseteq \mathcal{S}^\ell_\varnothing$$

for each $\ell \in \mathbb{N}$, where recall that by definition $\partial_+ \mathcal{S}^\ell_\varnothing$ is the vertical boundary of $\mathcal{S}_\varnothing$ restricted to its ℓth level $\mathcal{S}^\ell_\varnothing$. We say that the collar of the vertical boundary is *level preserving*.

The design piece corresponding to the empty word is defined to be

$$\mathfrak{D}_\varnothing := c_\varnothing\big(\partial_+ \mathcal{S}^1_\varnothing \times [0,1]\big) \cup c_-\big(\partial_- \mathcal{S}_\varnothing \times [0,1]\big).$$

In simple terms, $\mathfrak{D}_\varnothing$ is a neighbourhood of $\partial_- \widehat{\mathcal{S}}_\varnothing \cup \partial_+ \widehat{\mathcal{S}}^1_\varnothing$ within $\widehat{\mathcal{S}}_\varnothing$. However, the parametrisation of this neighbourhood, rather than merely the neighbourhood itself, will be important later. In what follows we iterate the production of collars as well as of skyscrapers.

28.3.2. Design pieces for finite binary words. Suppose that we have already built a skyscraper $\widehat{\mathcal{S}}_I \subseteq \widehat{\mathcal{S}}$ and a level preserving collar $c_I\colon \partial_+ \mathcal{S}_I \times [0,1] \to \mathcal{S}_I$ for a binary word I of finite length k. We build skyscrapers and collars for the binary words $I \cdot 0$ and $I \cdot 1$ as follows. Define

$$\widehat{\mathcal{S}}_{I\cdot 0} := \widehat{\mathcal{S}}_I.$$

Obtain the collar $c_{I\cdot 0}$ from c_I via the following modification.

$$\begin{aligned} c_{I\cdot 0}\colon \partial_+ \mathcal{S}_{I\cdot 0} \times [0,1] &\longrightarrow \mathcal{S}_{I\cdot 0} := \mathcal{S}_I \\ (x,t) &\longmapsto c_I(x, \tfrac{1}{3}t). \end{aligned}$$

Roughly speaking, we are using a third of the collar c_I of $\partial_+ \mathcal{S}_I = \partial_+ \mathcal{S}_{I\cdot 0}$. By construction,

$$c_{I\cdot 0}\big(\partial_+ \mathcal{S}^\ell_{I\cdot 0} \times [0,1]\big) \subseteq \mathcal{S}^\ell_{I\cdot 0}$$

for each $\ell \in \mathbb{N}$.

Obtaining $\widehat{\mathcal{S}}_{I\cdot 1}$ and the associated collar $c_{I\cdot 1}$ is more involved. Recall that $\mathcal{S}_I^{\leq \ell}$ denotes the levels of $\mathcal{S}_I$ up to and including level ℓ for each $\ell \in \mathbb{N}$. Recall also that $k = |I|$. In order to build $\widehat{\mathcal{S}}_{I\cdot 1}$, take $\widehat{\mathcal{S}}_I$ and discard everything above level $k+2$. Next, scrape off the two-thirds of the collar of $\partial_+\mathcal{S}_I$ defined by c_I closest to $\partial_+\mathcal{S}_I$. Note that this includes removing the part used in the construction of $c_{I\cdot 0}$. That is, consider the space

$$\widetilde{\mathcal{T}}_I := \mathcal{S}_I^{\leq k+2} \smallsetminus c_I\big(\partial_+\mathcal{S}_I \times \big[0, \tfrac{2}{3}\big)\big),$$

which is homeomorphic to $\mathcal{S}_I^{\leq k+2}$, and moreover, is a tower with $2k{+}4$ storeys (recall that a level of a skyscraper consists of two storeys), inheriting a decomposition into levels and blocks from $\mathcal{S}_I^{\leq k+2}$. Now zoom in on the two final storeys of $\widetilde{\mathcal{T}}_I$. Each connected component is a disc-like 2-storey tower (or more generally, in the terminology of Chapter 24, a room from a replicable class) attached to $\widetilde{\mathcal{T}}_I^{\leq 2k+2}$, the tower consisting of the first $2k+2$ storeys of $\widetilde{\mathcal{T}}_I$. The interior of each such 2-storey tower is an open subset of $\widehat{\mathcal{S}}$, so we can find within each such tower a closed metric ball of diameter strictly less than $1/(|I|+1)$, with positive distance from both the attaching region of that 2-storey tower and the image of the collar c_I. Call the union of these balls B_I. By the skyscraper embedding theorem (Theorem 26.10), we can find a skyscraper within each such 2-storey tower, with the same attaching region, such that the connected components of the second and higher levels of these new skyscrapers are contained in mutually disjoint balls B_I^j, where $B_I^j \subseteq B_I$, with positive pairwise distances between them. These newly found skyscrapers, attached to $\widetilde{\mathcal{T}}_I^{\leq 2k+2}$, yield the skyscraper that we call $\widehat{\mathcal{S}}_{I\cdot 1}$. The size and separation for the balls above will soon help us in building skyscrapers corresponding to infinite binary words, as well as in the proof of the β shrink (Section 28.6.3).

The collar $c_{I\cdot 1}$ is defined in two steps. The vertical boundary of the tower $\widetilde{\mathcal{T}}_I$ has a collar defined by the map $c_I|_{\partial_+\mathcal{S}_I \times [\frac{2}{3},1]}$. By the skyscraper embedding theorem, up to and including the first surface stage of the $(2k+3)$th storey, the towers $\widetilde{\mathcal{T}}_I$ and $\mathcal{S}_{I\cdot 1}^{\leq k+2}$ agree, so we take the collar of $\partial_+\mathcal{S}_{I\cdot 1}^{\leq k+2}$ to agree with $c_I|_{\partial_+\mathcal{S}_I \times [\frac{2}{3},1]}$. Thereafter, choose a parametrisation

$$c_{I\cdot 1} \colon \partial_+\mathcal{S}_{I\cdot 1} \times [0,1] \to \mathcal{S}_{I\cdot 1} \subseteq \widetilde{\mathcal{T}}_I$$

that extends the previously chosen parametrisation in the lower storeys. Once again, we require that

$$c_{I\cdot 1}\big(\partial_+\mathcal{S}_{I\cdot 1}^{\ell} \times [0,1]\big) \subseteq \mathcal{S}_{I\cdot 1}^{\ell}$$

for each $\ell \in \mathbb{N}$ i.e. that the collar be level preserving.

By induction, we can now construct skyscrapers and collars corresponding to every finite binary word. A schematic representation of the first few skyscrapers and collars are shown in Figures 28.3 and 28.4. Figure 28.5 shows these new embedded skyscrapers within the original skyscraper $\widehat{\mathcal{S}}$. Reconstructing these figures might help the reader understand the details of the construction.

The design piece $\mathfrak{D}_I$ associated to a finite binary word $I \in \{0,1\}^k$, with $k \neq 0$, is given by

$$\mathfrak{D}_I := \operatorname{Im} c_I|_{\partial_+\mathcal{S}_I^{k+1} \times [0,1]}.$$

This is a collar of $\partial_+\widehat{\mathcal{S}}_I$, restricted to the $(k+1)$th level. The shift by one matches the definition of $\mathfrak{D}_\varnothing$ above, which includes $c_\varnothing(\partial_+\mathcal{S}_\varnothing^1 \times [0,1])$, the collar of the first level. Levels of towers and skyscrapers use american numbering conventions for levels, where the ground floor is level 1.

This completes the construction of skyscrapers $\widehat{\mathcal{S}}_I \subseteq \widehat{\mathcal{S}}$, their collars c_I, and design pieces $\mathfrak{D}_I$ for all finite binary words I.

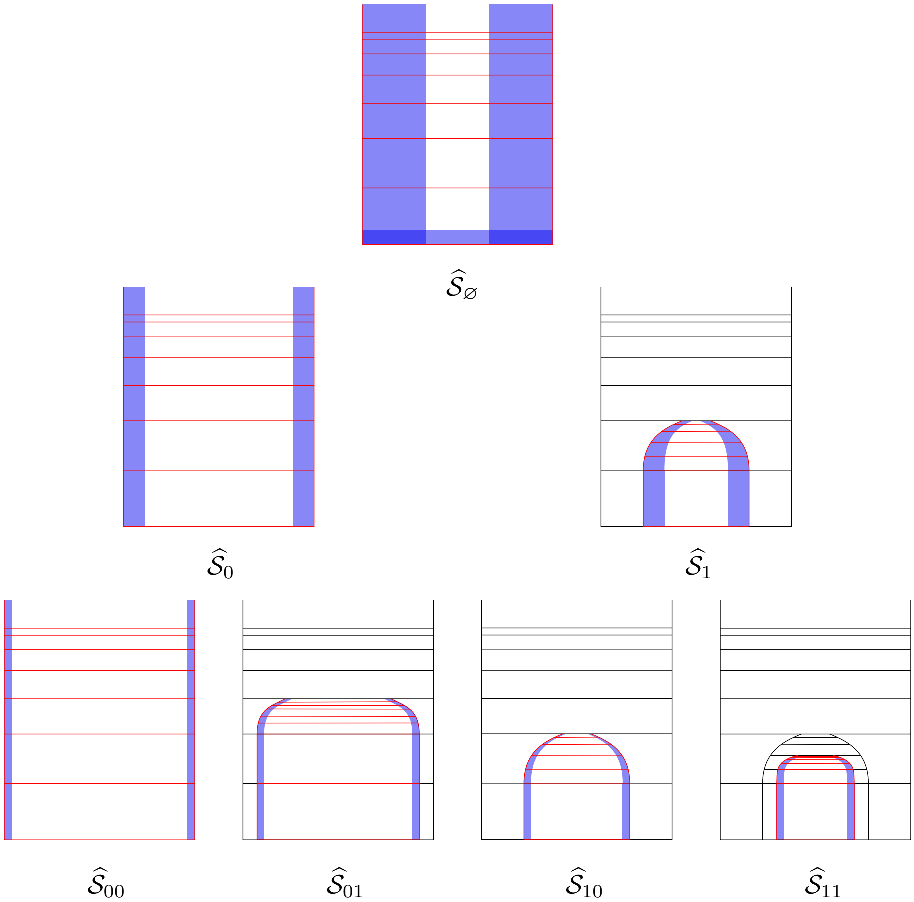


Figure 28.3. Skyscrapers and collars corresponding to finite binary words of length zero, one, and two, shown within the skyscraper $\widehat{\mathcal{S}}$. The skyscraper in question is shown in red, while previously used skyscrapers are in black. The corresponding collar is shown in blue. Note that the collars get thinner as the length of the corresponding finite binary word grows, and the attaching regions of the embedded skyscrapers are contained within the attaching region of $\widehat{\mathcal{S}}$.

We record the following property of the skyscrapers constructed above for use in the proof of Lemma 28.24.

Scholium 28.5. *Given a finite binary word I, if the kth digit is 1 for some $1 \leq k \leq |I|$, then the connected components of the closure of $\widehat{\mathcal{S}}_I \setminus \widehat{\mathcal{S}}_I^{\leq k+1}$ lie in mutually disjoint balls of diameter strictly less than $1/k$, and which furthermore have positive distance between one another.*

28.3.3. Design pieces for infinite binary words. Given an infinite binary word, each finite binary word obtained as a truncation has a corresponding skyscraper within $\widehat{\mathcal{S}}$ as constructed above. Intuitively, the skyscraper corresponding to the infinite binary word is the 'limit' of the skyscrapers corresponding to the truncations. We make this notion precise. Let $I = i_1 i_2 \dots$ be an infinite binary word. By our construction, we have that

$$\mathcal{S}_\varnothing \supsetneq \mathcal{S}_{i_1} \supsetneq \mathcal{S}_{i_1 i_2} \supsetneq \mathcal{S}_{i_1 i_2 i_3} \supsetneq \cdots$$

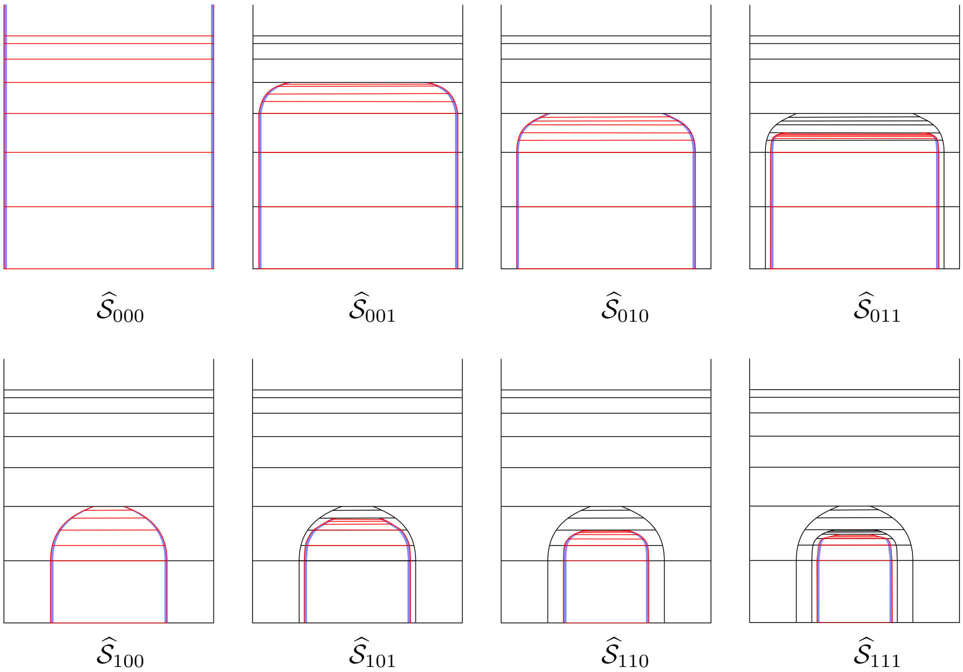


Figure 28.4. Skyscrapers and collars corresponding to finite binary words of length three, shown within the skyscraper $\widehat{\mathcal{S}}$. The skyscraper in question is shown in red, while previously used skyscrapers are in black. The collar is shown in blue.

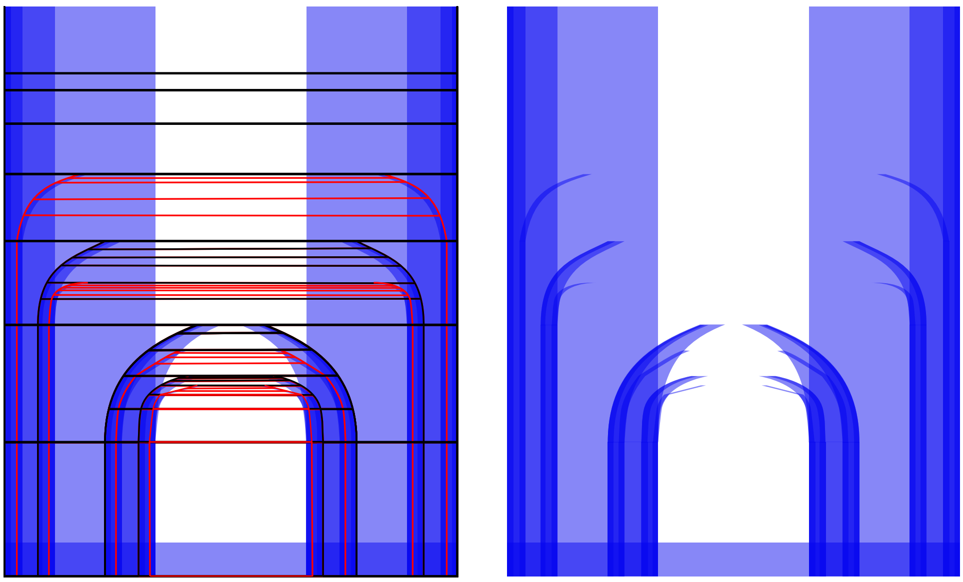


Figure 28.5. Skyscrapers and collars corresponding to finite binary words of length at most three, shown within the original skyscraper $\widehat{\mathcal{S}}$. In the right hand picture, only the collars are shown.

Let $\mathcal{S}_I$ denote the infinite intersection of this system of subsets. Again by construction, $\mathcal{S}_I$ is an open skyscraper: for each $j \geq 0$, the $(j+1)$th level of $\mathcal{S}_I$ is the

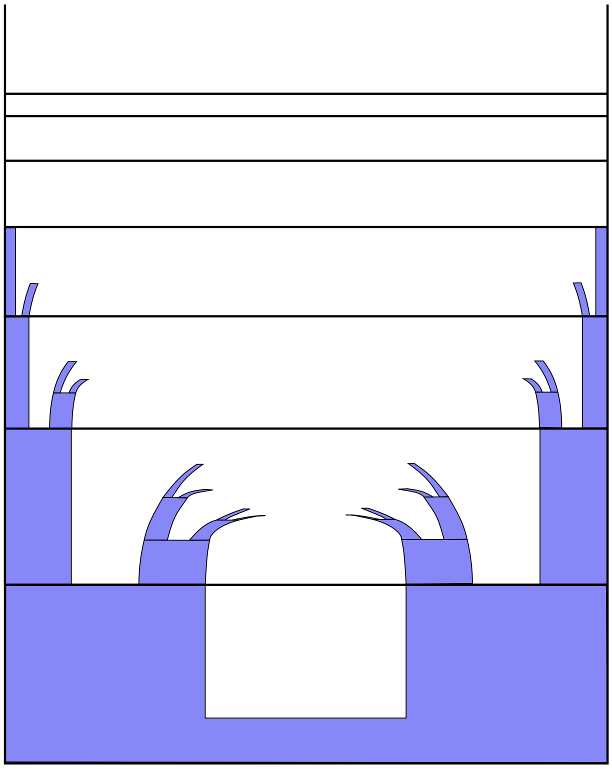


Figure 28.6. The design in the skyscraper $\widehat{\mathcal{S}}$. Only the levels for $\widehat{\mathcal{S}}$, rather than those for the embedded skyscrapers, are shown.

$(j+1)$th level of $\mathcal{S}_{i_1 i_2 \dots i_j}$, possibly with an open collar of the vertical boundary removed.

If I is essentially finite with an infinite tail of 0s, that is $I = I' \cdot 00\dots$ for some finite binary word I', then let

$$\mathcal{S}_I := \mathcal{S}_{I'} \text{ and } \widehat{\mathcal{S}}_I := \widehat{\mathcal{S}}_{I'}.$$

For other types of infinite binary words, that is those with infinitely many 1s, some more work will be needed.

Lemma 28.6. *Let I be an infinite binary word with infinitely many 1s. Let $\overline{\mathcal{S}}_I$ denote the closure of $\mathcal{S}_I$ in $\widehat{\mathcal{S}}$. Then $\overline{\mathcal{S}}_I$ is a skyscraper and in particular is homeomorphic to the endpoint compactification $\widehat{\mathcal{S}}_I$ of $\mathcal{S}_I$.*

In light of the lemma, we define

$$\widehat{\mathcal{S}}_I = \overline{\mathcal{S}}_I,$$

for an infinite binary word I and where $\overline{\mathcal{S}}_I$ denotes the closure of $\mathcal{S}_I$ in $\widehat{\mathcal{S}}$.

Proof of Lemma 28.6. As a closed subset of the compact space $\widehat{\mathcal{S}}$, the space $\overline{\mathcal{S}}_I$ is compact. We need to show that the limit points added to $\mathcal{S}_I$ to obtain $\overline{\mathcal{S}}_I$ can be identified with the ends of $\mathcal{S}_I$ and that the subspace topology on $\overline{\mathcal{S}}_I$ is the topology on the endpoint compactification $\widehat{\mathcal{S}}_I$ of $\mathcal{S}_I$.

The argument is similar to that in the proof of Theorem 18.9. Let x be a point in $\overline{\mathcal{S}}_I \smallsetminus \mathcal{S}_I$. Then by the definition of the closure of a set, the point x is the limit of a convergent sequence of points in $\mathcal{S}_I$, call it $\{a_j\}$. Each a_j is contained in some finite truncation $\mathcal{S}_I^{\leq k_j}$ of $\mathcal{S}_I$. Since x is not in $\mathcal{S}_I$, the sequence $\{k_j\}$ must be unbounded.

This follows from the fact that every finite truncation of $\mathcal{S}_I$ is a compact subset of the Hausdorff space $\widehat{\mathcal{S}}$ and is consequently closed and contains all its limit points. That is, the points a_j must eventually leave any finite truncation of $\mathcal{S}_I$. Let $\{a_{j_\ell}\}$ denote a subsequence of $\{a_j\}$ such that $\{k_{j_\ell}\}$ is strictly increasing. The sequence $\{a_{j_\ell}\}$ also converges to x. Let U_{j_ℓ} be the component of the complement of $\mathcal{S}_I^{\leq k_{j_\ell}}$ in $\mathcal{S}_I$ containing a_{j_ℓ}. Then we associate to x the sequence $(U_{j_1}, U_{j_2}, \dots)$. By the definition of ends of a space (Definition 12.17), such a sequence corresponds to a unique end of $\mathcal{S}_I$. (The sequence $(U_{j_1}, U_{j_2}, \dots)$ might not yet be an end of $\mathcal{S}_I$: in general we need to add some intervening open sets to obtain an end. However, the resulting end is uniquely determined.)

We need to show that this correspondence between limit points of $\overline{\mathcal{S}}_I$ contained in $\overline{\mathcal{S}}_I \setminus \mathcal{S}_I$ and the set of ends of $\mathcal{S}_I$ is a bijection. To do so we define a map in the other direction. Every end of $\mathcal{S}_I$ gives rise to a sequence of points, obtained by choosing points in components of complements of progressively taller finite truncations of $\mathcal{S}_I$. Since these components lie in metric balls of progressively smaller radii converging to zero, by the careful squeezing while embedding skyscrapers in Section 28.3.2, for each instance of 1 in the binary word I, the sequence converges. This is where we use the fact that there are infinitely many 1s in the binary word I. Since the sequence of points leaves any finite truncation of $\mathcal{S}_I$, the limit point must lie in $\overline{\mathcal{S}}_I \setminus \mathcal{S}_I$. Moreover, by the same argument, any two sequences corresponding to the same end have the same limit point. So we have defined a function Φ from the set of ends of $\mathcal{S}_I$ to the limit points of $\overline{\mathcal{S}}_I$ contained in $\overline{\mathcal{S}}_I \setminus \mathcal{S}_I$. In the squeezing process in Section 28.3.2, connected components of higher levels of an embedded skyscraper were placed in small balls with positive distance between them. Consequently, sequences of points associated with distinct ends of $\mathcal{S}_I$ cannot have the same limit point either. It follows that the function Φ is injective. The correspondence from the previous paragraph produces an end from a limit point, which by inspection maps back to the original limit point. Therefore Φ is surjective. This completes the proof that the set of ends of $\mathcal{S}_I$ is in bijective correspondence with the points of $\overline{\mathcal{S}}_I \setminus \mathcal{S}_I$.

It remains only to see that the topology on the endpoint compactification coincides with the subspace topology. By Definition 12.17, the endpoint compactification topology on $\widehat{\mathcal{S}}_I$ is generated by open sets of $\mathcal{S}_I$ and sets $V \subseteq \widehat{\mathcal{S}}_I$ such that $V \cap \mathcal{S}_I$ is a component U of the complement of some finite truncation $\widehat{\mathcal{S}}_I^{\leq k}$ of $\mathcal{S}_I$ and V contains the endpoints of $\mathcal{S}_I$ associated with U. The sets of the first type are open by construction, since $\mathcal{S}_I$ is constructed as a submanifold of $\mathcal{S}$. By the skyscraper embedding process in Section 28.3.2, sets of the second type are also open in the subspace topology. More precisely, in the skyscraper embedding process, for every instance of a 1 in the binary word I, the connected components of the complement of a finite truncation were embedded inside small balls, plus a collar on their attaching regions that stretches from the attaching regions to inside the small balls (see Figure 18.12). Take a union of the interior of one of these balls with an open neighbourhood of the collar to obtain an open subset of $\widehat{\mathcal{S}}$ whose intersection with $\overline{\mathcal{S}}_I$ is one of the generating open sets of the second type. This shows that the open sets in the endpoint compactification topology are open in the subspace topology.

Similarly, every open set in the subspace topology is open in the endpoint compactification topology. To see this, let U be an open set in $\widehat{\mathcal{S}}$ and consider the set $V := U \cap \overline{\mathcal{S}}_I$, which is open in the subspace topology by definition. A point in V is either an endpoint of $\widehat{\mathcal{S}}_I$ or contained in $\mathcal{S}_I$. If the latter, then it has a neighbourhood which is open in $\mathcal{S}_I$. If the former, then we claim the point has a neighbourhood which is a component of the complement of a finite truncation of $\widehat{\mathcal{S}}_I$ by construction. To see this claim, choose a suitably small ball $B_J^j \subseteq U$ in the construction of $\widehat{\mathcal{S}}_I$ in Section 28.3.2, for a sufficiently long finite subword $J \preceq I$, and then take a component of the complement of a finite truncation that is contained in that ball.

This completes the proof that $\overline{\mathcal{S}}_I$ is a skyscraper and that it is the endpoint compactification $\widehat{\mathcal{S}}_I$ of $\mathcal{S}_I$. $\square$

Next we define the design pieces for infinite binary words. Recall that as the length of a finite binary word grows, the collar of the corresponding skyscraper gets thinner within $\widehat{\mathcal{S}}$. In the limit, for the skyscrapers constructed above corresponding to infinite binary words, we do not define a collar at all. Rather, the design piece $\mathfrak{D}_I$ for an infinite binary word I is defined to be the collection of endpoints of the skyscraper $\widehat{\mathcal{S}}_I$. A word of caution: although the schematic pictures in Figures 28.3, 28.4, 28.5, and 28.6 seem to suggest that a skyscraper has a single end, in reality that is far from the truth.

Finally, define $\mathfrak{D} \subsetneq \widehat{\mathcal{S}}$ as the union of design pieces $\mathfrak{D}_I$ for all finite and infinite binary words, including $\varnothing$. That is,

$$\mathfrak{D} := \bigcup_{|I| \in \overline{\mathbb{N}}} \mathfrak{D}_I. \tag{28.2}$$

In the rest of this chapter, we will sometimes consider the union above as an abstract space, in which case it is simply called *the design*. Otherwise, we will consider the embedding of the design in $\widehat{\mathcal{S}}$, which is then called *the design in the skyscraper*.

Roughly speaking, the design in the skyscraper consists of collars of the vertical boundary for all of the embedded skyscrapers within $\widehat{\mathcal{S}}$, that taper towards (and include) the endpoints as we get higher (see Figure 28.6). The design piece $\mathfrak{D}_\varnothing$ corresponding to the empty word includes a collar for $\partial_-\widehat{\mathcal{S}}$. Note that we do not get a collar for each $\partial_+\widehat{\mathcal{S}}_I$, but only for $\partial_+\mathcal{S}_I$. Nonetheless, for every infinite binary word I, the entire vertical boundary $\partial_+\widehat{\mathcal{S}}_I$ of $\widehat{\mathcal{S}}_I$ is contained in the design. Since every finite binary word I can be completed to an (essentially finite) infinite binary word $I \cdot 000\ldots$, by construction the entire vertical boundary of every $\widehat{\mathcal{S}}_I$ is also contained in the design. Recall from Remark 26.6 that this vertical boundary is homeomorphic to $D^2 \times S^1$, which will be important later. Finally, the vertical boundary $\partial_+\widehat{\mathcal{S}}_\varnothing = \partial_+\widehat{\mathcal{S}}$ is also included in the design $\mathfrak{D}$.

We emphasise that we are in no way claiming that the design fills up all the space inside $\widehat{\mathcal{S}}$. The closures of the connected components of the complement of the design in $\widehat{\mathcal{S}}$ are called *gaps*. One could think of the design as an incomplete jigsaw puzzle for $\widehat{\mathcal{S}}$, where the gaps correspond to the missing pieces.

Let $\mathfrak{D}^{<\infty}$ denote the part of the design coming from finite binary words, and let $\mathfrak{D}^\infty$ be the part coming from infinite binary words. Here, by definition the infinite part $\mathfrak{D}^\infty$ consists of the endpoints of all the skyscrapers $\widehat{\mathcal{S}}_I$ with $|I|$ infinite. Since the endpoints of a given skyscraper form a Cantor set, $\mathfrak{D}^\infty$ is a "Cantor set's worth of Cantor sets" inside $\widehat{\mathcal{S}}$. Certainly, the finite part $\mathfrak{D}^{<\infty}$ seems more accessible than $\mathfrak{D}^\infty$. Fortuitously, we have the following scholium, which is directly implied by Lemma 28.6, the construction of embedded skyscrapers in Section 28.3.2 using the skyscraper embedding theorem (Theorem 26.10), and the fact that the ends of an open skyscraper are naturally identified with the ends of its vertical boundary (Remark 26.9).

SCHOLIUM 28.7. *The design $\mathfrak{D}$ in the skyscraper $\widehat{\mathcal{S}}$ is:*

(*a*) *the closure of $\mathfrak{D}^{<\infty}$ in $\widehat{\mathcal{S}}$; and*
(*b*) *homeomorphic to the endpoint compactification of $\mathfrak{D}^{<\infty}$.*

REMARK 28.8. The building instructions involved many choices, so it is hard to imagine that two implementations would yield the same set. We speak of *the* design only after these choices have been made. Any choice following the above recipe works for what comes next.

We finish this section by recording some properties of the construction so far. The following remark will be used in the construction of the design in the standard handle and the proof of Lemma 28.20.

REMARK 28.9. For binary words I and J, if $I \preceq J$, then $\widehat{\mathcal{S}}_J$ is a subset of $\mathcal{S}_I$ and either $\widehat{\mathcal{S}}_I^{|I|+1} = \widehat{\mathcal{S}}_J^{|I|+1}$ or $\widehat{\mathcal{S}}_J^{|I|+1}$ coincides with $\widehat{\mathcal{S}}_I^{|I|+1} \setminus c_I\big(\partial_+\mathcal{S}_I^{|I|+1} \times [0, \rho_J^I)\big)$ for some $\rho_J^I \in [0,1]$. By the level preserving nature of the collars, whenever $\widehat{\mathcal{S}}_J$ is contained in $\widehat{\mathcal{S}}_I$, and for each $j \leq |I|+1$, $\partial_+^{\leq j}\mathcal{S}_J$ meets the first j levels of $\widehat{\mathcal{S}}_I$ in $c_I\big(\partial_+\mathcal{S}_I^{\leq j} \times \{\rho_J^I\}\big)$ for $\rho_J^I \in [0,1]$ as above. This is visible in Figure 28.5.

The following lemma is a special case of the above remark for the skyscraper $\widehat{\mathcal{S}}_\varnothing = \widehat{\mathcal{S}}$, using the map $\Gamma\colon \mathcal{S} \hookrightarrow H$ from the beginning of Section 28.3.

LEMMA 28.10. *Let I be a binary word. The attaching region $\partial_-\widehat{\mathcal{S}}_I$ of the skyscraper $\widehat{\mathcal{S}}_I$ in $\widehat{\mathcal{S}}$ consists of the solid torus*

$$\partial_-\widehat{\mathcal{S}} \setminus c_\varnothing((\partial_+\widehat{\mathcal{S}} \cap \partial_-\widehat{\mathcal{S}}) \times [0, \mathfrak{c}_I))$$

which is mapped by $\Gamma\colon \mathcal{S} \hookrightarrow H$ to the solid torus

$$\{(1,\theta; r, \varphi) \mid 0 \leq r \leq 1 - \tfrac{\mathfrak{c}_I}{3}, 0 \leq \theta, \varphi \leq 2\pi\} \subseteq \partial_- H.$$

In particular, the torus $\partial_+\widehat{\mathcal{S}}_I \cap \partial_-\widehat{\mathcal{S}}_I$ coincides with

$$c_\varnothing((\partial_+\widehat{\mathcal{S}} \cap \partial_-\widehat{\mathcal{S}}) \times \{\mathfrak{c}_I\})$$

and is mapped by Γ to

$$\{(1,\theta; 1 - \tfrac{\mathfrak{c}_I}{3}, \varphi) \mid 0 \leq \theta, \varphi \leq 2\pi\} \subseteq \partial_- H.$$

PROOF. That the attaching region of $\widehat{\mathcal{S}}_I$ consists of the solid torus $\partial_-\widehat{\mathcal{S}} \setminus c_\varnothing((\partial_+\widehat{\mathcal{S}} \cap \partial_-\widehat{\mathcal{S}}) \times [0, \mathfrak{c}_I))$ follows immediately from the construction of $\widehat{\mathcal{S}}_I$. For the correspondence under Γ, recall that we defined $c_\varnothing(x,t) = \Gamma^{-1}(C_+(\Gamma(x),t))$ for any $x \in \partial_+\mathcal{S}_\varnothing = \partial_+\mathcal{S}$ and $t \in [0,1]$, where C_+ is the collar of $\partial_+ H$ from Section 28.2. In our case $x \in \partial_-\widehat{\mathcal{S}} \cap \partial_+\widehat{\mathcal{S}}$. Recall also that $\Gamma(\partial_-\widehat{\mathcal{S}} \cap \partial_+\widehat{\mathcal{S}}) = \partial_- H \cap \partial_+ H$. Then $\Gamma(c_\varnothing(x,t)) = C_+(\Gamma(x),t)$, where $\Gamma(x) \in \partial_- H \cap \partial_+ H$ so is of the form $(1,\theta;1,\varphi)$ for some $0 \leq \theta, \varphi \leq 2\pi$. By the definition of C_+, we have $C_+(\Gamma(x),t) = (1,\theta; 1-\frac{t}{3}, \varphi)$, as needed. $\square$

We will also need the following relation between the various embedded skyscrapers $\widehat{\mathcal{S}}_I$ in the proof of Lemma 28.20.

LEMMA 28.11. *For binary words I and J, we have $\mathfrak{c}_I < \mathfrak{c}_J$ if and only if $\widehat{\mathcal{S}}_J \subsetneq \widehat{\mathcal{S}}_I$, in which case $\partial_+\widehat{\mathcal{S}}_J$ separates $\widehat{\mathcal{S}}_I$.*

PROOF. The first part follows from the construction (see Figure 28.5). To be more precise, if $\widehat{\mathcal{S}}_J \subsetneq \widehat{\mathcal{S}}_I$ then $\mathfrak{c}_I < \mathfrak{c}_J$ by Lemma 28.10. For the other direction, assume $\mathfrak{c}_I < \mathfrak{c}_J$. Without loss of generality, assume that both I and J are infinite binary words, by adding an infinite tail of 0s if necessary. Let i_ℓ and j_ℓ denote the ℓth digit of I and J respectively, for each ℓ. Let n be the least integer such that $i_n \neq j_n$. We claim that $i_n = 0$ and $j_n = 1$. To see the claim note that the alternative is that $i_n = 1$ and $j_n = 0$, in which case

$$\mathfrak{c}_J \leq \sum_{\ell=0}^{n-1} \frac{2j_\ell}{3^\ell} + \sum_{\ell=n+1}^{\infty} \frac{2}{3^\ell} = \sum_{\ell=0}^{n-1} \frac{2j_\ell}{3^\ell} + \frac{1}{3^n}$$

and so

$$\mathfrak{c}_I = \sum_{\ell=0}^{n-1} \frac{2j_\ell}{3^\ell} + \frac{2}{3^n} + \sum_{\ell=n+1}^{\infty} \frac{2i_\ell}{3^\ell} > \sum_{\ell=0}^{n-1} \frac{2j_\ell}{3^\ell} + \frac{1}{3^n} \geq \mathfrak{c}_J,$$

which is a contradiction. This shows that $i_n = 0$ and $j_n = 1$. Then by the construction in Section 28.3.2, we have that $\widehat{\mathcal{S}}_J \subsetneq \widehat{\mathcal{S}}^{\leq n+1}_{i_1 i_2 \dots i_{n-1} 0} \subsetneq \widehat{\mathcal{S}}_I$, as needed.

It remains to prove that if $\widehat{\mathcal{S}}_J \subsetneq \widehat{\mathcal{S}}_I$, then $\partial_+ \widehat{\mathcal{S}}_J$ separates $\widehat{\mathcal{S}}_I$, in other words that $\widehat{\mathcal{S}}_I \smallsetminus \partial_+ \widehat{\mathcal{S}}_J$ is disconnected. Note that every skyscraper is locally path connected as a quotient of a locally path connected space (Proposition 14.1). Then the open subset $\widehat{\mathcal{S}}_I \smallsetminus \partial_+ \widehat{\mathcal{S}}_J \subseteq \widehat{\mathcal{S}}_I$ is also locally path connected, so it is connected if and only if it is path connected.

We will show that $\widehat{\mathcal{S}}_I \smallsetminus \partial_+ \widehat{\mathcal{S}}_J$ is not path connected. Suppose for a contradiction that it is. Then there is a path γ joining a point $p \in \widehat{\mathcal{S}}_J \smallsetminus \partial_+ \widehat{\mathcal{S}}_J$ and a point q in $\widehat{\mathcal{S}}_I \smallsetminus \widehat{\mathcal{S}}_J$ such that $\gamma \cap \partial_+ \widehat{\mathcal{S}}_J = \emptyset$. Since $\widehat{\mathcal{S}}_I$ is closed in $\widehat{\mathcal{S}}$, we know that $\gamma \cap \widehat{\mathcal{S}}_J$ is closed in $\widehat{\mathcal{S}}$.

Now we argue that $\gamma \cap \widehat{\mathcal{S}}_J$ cannot be contained in any finite truncation of $\widehat{\mathcal{S}}_J$. To see this, we assert that the boundary of every finite truncation $\widehat{\mathcal{S}}^{\leq k}_J$ of $\widehat{\mathcal{S}}_J$ separates $\widehat{\mathcal{S}}_I$. To show the assertion, note that the vertical boundary $\partial_+ \widehat{\mathcal{S}}^{\leq k}_J$ is bicollared in $\widehat{\mathcal{S}}_I$, and moreover is contained in some finite truncation $\widehat{\mathcal{S}}^{\leq \ell-1}_I$ of $\widehat{\mathcal{S}}_I$. We apply the Mayer-Vietoris sequence using the decomposition $\widehat{\mathcal{S}}^{\leq \ell}_I = X \cup Y$, where $X := \mathrm{cl}(\nu \partial_+ \widehat{\mathcal{S}}^{\leq k}_J)$ is the bicollar mentioned above, $Y = \widehat{\mathcal{S}}^{\leq \ell}_I \smallsetminus \partial_+ \widehat{\mathcal{S}}^{\leq k}_J$, and their intersection is homotopy equivalent to two copies of X. Then since $\partial_+ \widehat{\mathcal{S}}^{\leq k}_J$ is contained in $\widehat{\mathcal{S}}^{\leq \ell-1}_I$, and the fundamental group of $\widehat{\mathcal{S}}^{\leq \ell}_I$ is generated by double point loops on the top stage caps, the homomorphism

$$H_1(Y;\mathbb{Z}) \to H_1(X \cup Y;\mathbb{Z})$$

is surjective. (In the language of Chapter 24, the double point loops should be replaced by the meridians of the link L_R.) The Mayer-Vietoris sequence then reduces to

$$\xrightarrow{0} H_0(X \sqcup X;\mathbb{Z}) \to H_0(X;\mathbb{Z}) \oplus H_0(Y;\mathbb{Z}) \to H_0(X \cup Y;\mathbb{Z}) \to 0.$$

Since $H_0(X;\mathbb{Z}) \cong H_0(X \cup Y;\mathbb{Z}) \cong \mathbb{Z}$, we obtain the short exact sequence

$$0 \to \mathbb{Z} \oplus \mathbb{Z} \to H_0(Y;\mathbb{Z}) \oplus \mathbb{Z} \to \mathbb{Z} \to 0,$$

from which it follows that $H_0(Y;\mathbb{Z}) \cong \mathbb{Z} \oplus \mathbb{Z}$. Therefore $\partial_+ \widehat{\mathcal{S}}^{\leq k}_J$ separates $\widehat{\mathcal{S}}^{\leq \ell}_I$. Then note that $\widehat{\mathcal{S}}_I$ is obtained from $\widehat{\mathcal{S}}^{\leq \ell}_I$ by gluing on skyscrapers, that are separated from $\widehat{\mathcal{S}}^{\leq \ell-1}_I$, along their attaching regions. Since $\partial_+ \widehat{\mathcal{S}}^{\leq k}_J \subseteq \widehat{\mathcal{S}}^{\leq \ell-1}_I$, this implies that all of the attaching regions lie in one connected component of $Y = \widehat{\mathcal{S}}^{\leq \ell}_I \smallsetminus \partial_+ \widehat{\mathcal{S}}^{\leq k}_J$. and thus $\partial_+ \widehat{\mathcal{S}}^{\leq k}_J$ also separates $\widehat{\mathcal{S}}_I$, as asserted.

It follows that $\gamma \cap \widehat{\mathcal{S}}_J$ contains points in arbitrarily tall truncations of $\widehat{\mathcal{S}}_J$. The limit of (a convergent subsequence of) these points is an endpoint of $\mathcal{S}_J$, so in particular an element of $\partial_+ \widehat{\mathcal{S}}_J$. Since the closed set $\gamma \cap \widehat{\mathcal{S}}_J$ contains all its limit points by definition, $\gamma \cap \partial_+ \widehat{\mathcal{S}}_J \neq \emptyset$, which is a contradiction. This completes the proof. $\square$

28.4. Embedding the design in the standard handle

Next we embed the design in the standard handle $H := D^2 \times D^2$. Once again we will use the explicit parametrisation from Section 28.2, using polar coordinates in each disc factor. In other words we write $H = \{(s,\theta; r,\varphi)\}$, where $0 \leq r, s \leq 1$ and $0 \leq \theta, \varphi \leq 2\pi$. We use the notation $\partial_- H := S^1 \times D^2 = \{s = 1\}$ and

$\partial_+ H := D^2 \times S^1 = \{r = 1\}$, as well as the notation $D^2 \times S^1_r$ for the solid torus $\{(s, \theta; r, \varphi\}_{s,\theta,\varphi}$ for fixed r.

28.4.1. Embedding finite word design pieces in H. We focus first on the finite part of the design $\mathfrak{D}^{<\infty} = \bigcup_{|I|<\infty} \mathfrak{D}_I$. Let I be a finite binary word with $|I| = k$. By definition, we have

$$\mathfrak{D}_\varnothing := c_\varnothing\big(\partial_+ \mathcal{S}^1_\varnothing \times [0,1]\big) \cup c_-\big(\partial_- \mathcal{S}_\varnothing \times [0,1]\big),$$

and if $k \neq 0$, then

$$\mathfrak{D}_I := \operatorname{Im} c_I\big|_{\partial_+ \mathcal{S}^{k+1}_I \times [0,1]}.$$

That is, for $k \neq 0$, (restrictions of) the collars give diffeomorphisms

$$\partial_+ \mathcal{S}^{k+1}_I \times [0,1] \xrightarrow[\cong]{c_I} \mathfrak{D}_I \subseteq \widehat{\mathcal{S}}. \tag{28.3}$$

We also saw in Remark 28.9 that if J is a binary word such that $I \preceq J$, then $\widehat{\mathcal{S}}_J$ is contained in $\widehat{\mathcal{S}}_I$, and for $j \leq k+1$, $\partial_+^{\leq j} \mathcal{S}_J$ meets the first j levels of $\widehat{\mathcal{S}}_I$ in $c_I\big(\partial_+ \mathcal{S}^{\leq j}_I \times \{\rho^I_J\}\big)$ for some $\rho^I_J \in [0,1]$. This gives us identifications, for every $j \leq k+1$ and for every $I \preceq J$,

$$\partial_+ \mathcal{S}^{\leq j}_I \xrightarrow{\equiv} \partial_+ \mathcal{S}^{\leq j}_J \tag{28.4}$$

induced by the collars.

In order to proceed, we need to know what these pieces $\partial_+^{\leq j} \widehat{\mathcal{S}}_I$ are. This was addressed in Remark 26.6, where we saw that for an open skyscraper $\mathcal{S}'$ and for every $j \in \mathbb{N}$, $\partial_+(\mathcal{S}')^{\leq j}$ is diffeomorphic to the complement of a tubular neighbourhood of a mixed ramified Bing-Whitehead link in $D^2 \times S^1$ with ramified Whitehead doubles in its last stage.

Then, for $0 \leq |I| < \infty$, we choose embeddings

$$\psi_I \colon \partial_+^{\leq k+1} \widehat{\mathcal{S}}_I \hookrightarrow D^2 \times S^1 \tag{28.5}$$

with the following properties. Warning: we will modify this choice of the ψ_I shortly, in Lemma 28.14, while preserving these properties.

(H1) The embedding $\psi_\varnothing$ is defined to be the composition

$$\partial^1_+ \widehat{\mathcal{S}} \hookrightarrow \partial_+ \mathcal{S} \overset{\Gamma}{\hookrightarrow} \partial_+ H = D^2 \times S^1,$$

for the map $\Gamma \colon \mathcal{S} \hookrightarrow H$ from the beginning of Section 28.3.

(H2) For $0 < |I|$, the embedding ψ_I restricted to the torus $\partial_+ \widehat{\mathcal{S}}_I \cap \partial_- \widehat{\mathcal{S}}_I$ is the composition

$$\partial_+ \widehat{\mathcal{S}}_I \cap \partial_- \widehat{\mathcal{S}}_I \xrightarrow[\cong]{\Gamma} \{(1, \theta; 1 - \tfrac{\mathfrak{c}_I}{3}, \varphi)\}_{\theta,\varphi} \hookrightarrow \{(s, \theta; 1 - \tfrac{\mathfrak{c}_I}{3}, \varphi)\}_{s,\theta,\varphi} =: D^2 \times S^1_{1-\frac{\mathfrak{c}_I}{3}}$$

followed by the canonical identification $D^2 \times S^1_{1-\frac{\mathfrak{c}_I}{3}} = D^2 \times S^1$ (see Lemma 28.10).

(H3) Whenever $I \preceq J$, we have a commutative diagram

$$\begin{tikzcd} \partial_+ \mathcal{S}^{\leq k+1}_I \arrow[r, hook, "\psi_I"] \arrow[d, "\equiv"] & D^2 \times S^1 \\ \partial_+ \widehat{\mathcal{S}}^{\leq k+1}_J \arrow[d, hook] & \\ \partial_+ \mathcal{S}^{\leq \ell+1}_J \arrow[uur, hook, "\psi_J"'] & \end{tikzcd} \tag{28.6}$$

where $|J| = \ell \geq k = |I|$. In the language of Chapter 13 and Kirby diagrams (see Remark 26.6), this means that the mixed ramified Bing-Whitehead links for $\partial_-\widehat{\mathcal{S}}_I^{\leq k+1}$ and $\partial_-\widehat{\mathcal{S}}_J^{\leq \ell+1}$ are drawn compatibly.

For a finite binary word $I = i_1 i_2 \ldots i_k$, define a rescaling function

$$\begin{aligned} r_I\colon [0,1] &\longrightarrow [0,1] \\ t &\longmapsto 1 - \frac{1}{3} a_I(t) \\ &= 1 - \frac{1}{3}\Big(\sum_{j=1}^{k} \frac{2 i_j}{3^j} + \frac{1}{3^k} t\Big). \end{aligned}$$

Recall that C_I is an interval *not* removed in the $|I|$th stage of constructing the ternary Cantor set on $[0,1]$ and a_I is the affine parametrisation of C_I from Section 28.1. The functions r_I affinely parametrise (while reversing the direction) the subintervals of $[\frac{2}{3},1]$ that are not removed in the $|I|$th step of constructing the ternary Cantor set on $[\frac{2}{3},1]$. For example, when $k=0$ and I is empty, we get $r_I(t) = 1 - \frac{t}{3}$. When $I = 0$, $k = 1$, we have $r_I(t) = 1 - \frac{t}{9}$. When $I = 1$, $k = 1$, we have $r_I(t) = \frac{7}{9} - \frac{t}{9}$.

Now we begin to define the embedding of the design in H, starting with the design piece for the empty word. Recall that

$$\mathfrak{D}_\varnothing := c_\varnothing\big(\partial_+\mathcal{S}^1_\varnothing \times [0,1]\big) \cup c_-\big(\partial_-\mathcal{S}_\varnothing \times [0,1]\big),$$

where $\mathcal{S}_\varnothing = \mathcal{S}$ by definition. Recall also that $c_-(x,t) = \Gamma^{-1}(C_-(\Gamma(x),t))$ for all $x \in \partial_-\mathcal{S}_\varnothing$ and $t \in [0,1]$, and similarly $c_\varnothing(x,t) = \Gamma^{-1}(C_+(\Gamma(x),t))$ for any $x \in \partial_+\mathcal{S}_\varnothing$ and $t \in [0,1]$.

We will define a function $\Psi_\varnothing$ on the two pieces separately and then confirm that the definition agrees on the overlap. This is the content of the following lemma.

Lemma 28.12. *Define*

$$\begin{aligned} \Psi_\varnothing\colon c_-\big(\partial_-\mathcal{S}_\varnothing \times [0,1]\big) &\to C_-(\partial_- H \times [0,1]) \subseteq H \\ \Gamma^{-1}\big(C_-(\Gamma(x),t)\big) &\mapsto C_-(\Gamma(x),t) \end{aligned}$$

for $x \in \partial_-\mathcal{S}_\varnothing$ and $t \in [0,1]$, and similarly

$$\begin{aligned} \Psi_\varnothing\colon c_\varnothing\big(\partial_+\mathcal{S}^1_\varnothing \times [0,1]\big) &\to C_+(\partial_+ H \times [0,1]) \subseteq H \\ \Gamma^{-1}\big(C_+(\Gamma(x),t)\big) &\mapsto C_+(\Gamma(x),t) \end{aligned}$$

for $x \in \partial_+\mathcal{S}^1_\varnothing$ and $t \in [0,1]$. Together these give a well defined function

$$\Psi_\varnothing\colon \mathfrak{D}_\varnothing \to H,$$

restricting to a diffeomorphism

$$\Psi_\varnothing|_{c_-\left(\partial_-\mathcal{S}_\varnothing \times [0,1]\right)}\colon c_-\big(\partial_-\mathcal{S}_\varnothing \times [0,1]\big) \to C_-(\partial_- H \times [0,1])$$

and so that $\Psi_\varnothing|_{\partial_-\mathcal{S}_\varnothing} = \Gamma|_{\partial_-\mathcal{S}_\varnothing}\colon \partial_-\mathcal{S}_\varnothing \to \partial_- H$.

Proof. Suppose there are points $x_\pm \in \partial_\pm\mathcal{S}$ and $t_\pm \in [0,1]$ with $\Gamma^{-1}(C_-(\Gamma(x_-),t_-)) = \Gamma^{-1}(C_+(\Gamma(x_+),t_+))$. We check that the two definitions of $\Psi_\varnothing$ agree, that is $C_-(\Gamma(x_-),t_-) = C_+(\Gamma(x_+),t_+)$.

Since Γ is well defined, we have that $C_-(\Gamma(x_-),t_-) = C_+(\Gamma(x_+),t_+)$. Denote $\Gamma(x_\pm)$ by $p_\pm$, where $p_- := (1,\theta_-;r_-,\varphi_-)$ and $p_+ := (s_+,\theta_+;1,\varphi_+)$, for some $r_-, s_+ \in [0,1]$ and $\theta_\pm, \varphi_\pm \in [0,2\pi]$. Then by Lemma 28.4, we have $\theta_- = \theta_+$, $\varphi_- = \varphi_+$, $1 - t_-/3 = s_+$, and $r_- = 1 - t_+/3$. By definition we have

$$C_-(\Gamma(x_-),t_-) = C_-((1,\theta_-;r_-,\varphi_-),t_-) = (1 - \tfrac{t_-}{3}, \theta_-; r_-, \varphi_-)$$

and

$$C_+(\Gamma(x_+), t_+) = C_+((s_+, \theta_+; 1, \varphi_+), t_+) = (s_+, \theta_+; 1 - \tfrac{t_+}{3}, \varphi_+)$$

which are equal, as needed.

The final restriction property follows since both collars C_- and c_- restrict to the identity for $0 \in [0,1]$. In other words, write $x \in \partial_-\mathcal{S}_\varnothing$ as $\Gamma^{-1}(C_-(\Gamma(x), 0))$. Then $\Psi_\varnothing(x) = C_-(\Gamma(x), 0) = \Gamma(x)$. $\square$

Note in particular that by Lemma 28.12, the map $\Psi_\varnothing$ sends $c_\varnothing\big(\partial_+\mathcal{S}^1_\varnothing \times [0,1]\big)$ to within $D^2 \times (S^1 \times [\frac{2}{3}, 1]) \subseteq D^2 \times D^2 = H$, and the region $c_-\big(\partial_-\mathcal{S}_\varnothing \times [0,1]\big)$ onto $\big(S^1 \times [\frac{2}{3}, 1]\big) \times D^2 \subseteq D^2 \times D^2 = H$.

Next, for $k = |I| \neq 0$, we define embeddings $\Psi_I \colon \mathfrak{D}_I \hookrightarrow D^2 \times D^2$ as the composition

$$\mathfrak{D}_I \xrightarrow[\cong]{c_I^{-1}} \partial_+\mathcal{S}_I^{k+1} \times [0,1] \xhookrightarrow{\psi_I \times r_I} D^2 \times S^1 \times [r_I(1), r_I(0)] \hookrightarrow D^2 \times D^2$$

where the last arrow comes from using polar coordinates in the second factor of $D^2 \times D^2$.

By construction, the embeddings defined above are such that the following lemma holds.

Lemma 28.13. *For any two finite binary words I and J, the maps Ψ_I and Ψ_J coincide on $\mathfrak{D}_I \cap \mathfrak{D}_J$, and thus we have a well defined embedding*

$$\Psi^{<\infty} := \bigcup_{|I|<\infty} \Psi_I \colon \mathfrak{D}^{<\infty} \to D^2 \times D^2 =: H,$$

restricting to the identity on $\partial_-\widehat{\mathcal{S}} \subseteq \mathfrak{D}_\varnothing$.

We give a schematic of the image of $\Psi_I^{<\infty}$ in Figure 28.7, using the same conventions as in Figure 28.2. For a finite binary word I, the radius $r = r_I(\frac{1}{2})$ corresponds

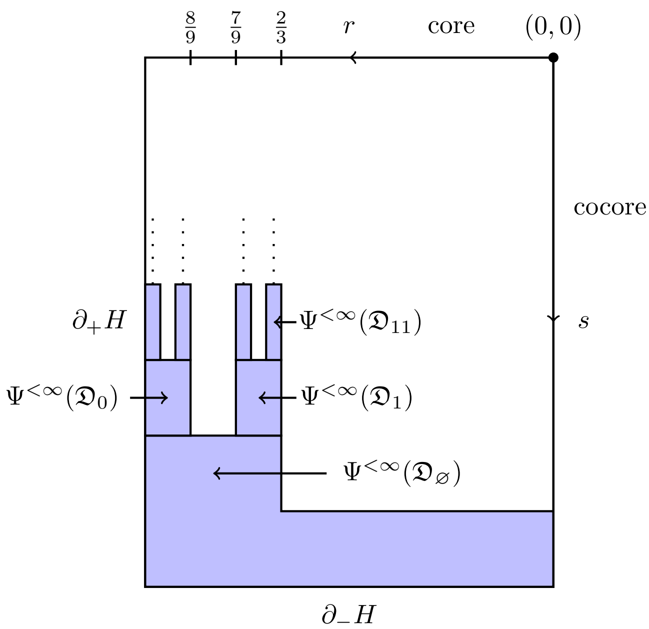


Figure 28.7. The design in the standard handle H.

to a vertical line in the figure, which describes a solid torus $D^2 \times S^1_r$. This is part of a thickened solid torus

$$D^2 \times S^1 \times (r_{I\cdot 1}(0), r_{I\cdot 0}(1)),$$

where the latter coordinate gives the radius. By definition, in each such $D^2 \times S^1_r$, the contribution to the design is the complement of a mixed ramified Bing-Whitehead link. Thus the subset that is not in the design in $D^2 \times S^1_r$ is a tubular neighbourhood of such a link. The white regions in the figure are the products of such neighbourhoods with $(r_{I\cdot 1}(0), r_{I\cdot 0}(1))$, within the thickened solid tori $D^2 \times S^1 \times (r_{I\cdot 1}(0), r_{I\cdot 0}(1))$.

Before we consider infinite words, a remark on terminology. We will refer to the image of the design in H as *the design in H* or *the design in the standard handle*, when it is necessary to differentiate from the design in the skyscraper.

28.4.2. Embedding infinite word design pieces in H. Let $I = i_1 i_2 i_3 \dots$ be an infinite binary word. Recall that $\mathfrak{D}_I$ is defined to be the collection of endpoints of the skyscraper $\widehat{\mathcal{S}}_I$. Let $r = 1 - \frac{1}{3}\Sigma_{j=1}^{\infty} \frac{2i_j}{3^j}$. It is straightforward to check that

$$r \in \cdots \subsetneq [r_{i_1 i_2 i_3}(1), r_{i_1 i_2 i_3}(0)] \subsetneq [r_{i_1 i_2}(1), r_{i_1 i_2}(0)] \subsetneq [r_{i_1}(1), r_{i_1}(0)].$$

Within the solid torus $D^2 \times S^1_r$, the contribution to the design for each such finite truncation of I is the complement of a tubular neighbourhood of a mixed ramified Bing-Whitehead link. Moreover, as we saw in Remark 26.6, for any two consecutive truncations the neighbourhoods are nested. So when we consider all such finite truncations, the contribution to the design in $D^2 \times S^1_r$ is the complement of the infinite intersection of these neighbourhoods. This is reminiscent of our work with decompositions in Part I, and this reminiscence is not a coincidence. The plan is to modify the embeddings for skyscrapers corresponding to finite binary words in the previous section, so that the infinite intersection of the tubular neighbourhoods of the corresponding mixed ramified Bing-Whitehead links is precisely the set of endpoints of the skyscrapers corresponding to infinite words. Thus when we embed the design pieces corresponding to infinite binary words in H, we are just adding these points back. Consequently, the entire solid torus $D^2 \times S^1_r$ will be contained in the design in H. This is the content of the following lemma.

LEMMA 28.14 (Automatic shrinking). *Let $I = i_1 i_2 \dots$ be an infinite binary word. Then the components of the complement of $\bigcup_k \operatorname{Im}\psi_{i_1\dots i_k} \subseteq D^2 \times S^1$ form a mixed ramified Bing-Whitehead decomposition $\mathcal{B}_I$ whose elements correspond bijectively to the endpoints of $\widehat{\mathcal{S}}_I$. Furthermore, the ψ maps in (28.5) can be chosen so that all the decomposition elements of $\mathcal{B}_I$ are points. Finally, for distinct infinite binary words I and J, the defining sequences for the corresponding mixed ramified Bing-Whitehead decompositions may be chosen compatibly.*

PROOF. According to Remark 26.6, the complement of $\operatorname{Im}\psi_{i_1\dots i_k}$ is a tubular neighbourhood of a mixed ramified Bing-Whitehead link. Denote this neighbourhood by L^k_I. By the compatibility condition (H3) we see that L^{k+1}_I is contained in L^k_I for each k so that the connected components of the intersection $\bigcap_k L^k_I$, which is the set of connected components of the complement of $\bigcup_k \operatorname{Im}\psi_{i_1\dots i_k}$, form a mixed ramified Bing-Whitehead decomposition $\mathcal{B}_I$ of $D^2 \times S^1$.

In order to relate this decomposition to $\widehat{\mathcal{S}}_I$, note that by construction $\widehat{\mathcal{S}}_I$ is contained in $\widehat{\mathcal{S}}_{i_1\dots i_k}$ for each k and the collar $c_{i_1\dots i_k}$ gives an identification

$$\partial^{\leq k+1}_{+} \widehat{\mathcal{S}}_{i_1\dots i_k} \longrightarrow \partial^{\leq k+1}_{+} \widehat{\mathcal{S}}_I$$

as in (28.4). This gives a diffeomorphism

$$\partial_+\mathcal{S}_I \xrightarrow{\cong} \bigcup_k \operatorname{Im}\psi_{i_1\dots i_k} \subseteq D^2 \times S^1.$$

Thus the ends of $\partial_+\mathcal{S}_r$ correspond bijectively with those of $\bigcup_k \operatorname{Im}\psi_{i_1\dots i_k}$, which in turn correspond to the elements of $\mathcal{B}_I$ by construction. On the other hand, one readily checks that the ends of $\partial_+\widehat{\mathcal{S}}_I$ correspond to those of $\widehat{\mathcal{S}}_I$ (see Remark 26.9).

The argument above gives an identification of $\partial_+\widehat{\mathcal{S}}_I$ with the decomposition space $D^2 \times S^1/\mathcal{B}_I$, and by the boundary shrinkable property of a skyscraper, the decomposition $\mathcal{B}_I$ is shrinkable.

By Scholium 26.2, the defining sequence $\{L_I^k\}_k$ of $\mathcal{B}_I$ can be repositioned by successively applying isotopies in deeper stages so that all the decomposition elements of $\mathcal{B}_I$ are points. This repositioning was already performed when defining the map $\Gamma\colon \mathcal{S} \hookrightarrow H$ at the start of Section 28.3 (see condition (S2)). Moreover, if two infinite binary words I, J agree in their first N digits, then for $k \leq N$ we can identify L_I^k and L_J^k according to (H3), and the shrinking isotopies can be chosen compatibly. Finally, these modifications of the $\{L_I^k\}_k$ can be translated into modifications of the ψ maps for finite binary words. $\square$

If the ψ maps are chosen as in Lemma 28.14, then we say that the embedding of the design in H has *automatic shrinking*. We will assume this property from now on. Automatic shrinking shows that the closure of $\bigcup_{|I|<\infty} \Psi_I(\mathfrak{D}^{<\infty})$ inside $D^2 \times D^2$ can be identified with the endpoint compactification of $\mathfrak{D}^{<\infty}$. According to Scholium 28.7 and the discussion preceding it, $\bigcup_{|I|<\infty} \Psi_I$ extends to an embedding of the full design

$$\Psi\colon \mathfrak{D} \to D^2 \times D^2 = H. \tag{28.7}$$

Here, for an infinite binary word $I = i_1 i_2 \dots$ and $r = 1 - \frac{1}{3}\Sigma_{j=1}^{\infty} \frac{2i_j}{3^j}$ as before, we send $\mathfrak{D}_I$, consisting of the endpoints of $\widehat{\mathcal{S}}_I$, to the complement of the set $\bigcup_k \operatorname{Im}\psi_{i_1\dots i_k} \times \{r\} \subseteq D^2 \times S^1 \times \{r\} = D^2 \times S^1_r$. The subset $\Psi(\mathfrak{D}) \subsetneq H$ is denoted D. The closures of the connected components of the complement of D in H are called *holes*.

Since Ψ is a topological embedding, it defines a homeomorphism $\mathfrak{D} \cong D$ that we will make use of later. We record the following property of the map Ψ.

LEMMA 28.15. *The map $\Psi\colon \mathfrak{D} \to D$ restricts to a diffeomorphism*

$$\Psi|_{c_-\left(\partial_-\widehat{\mathcal{S}}\times[0,1]\right)}\colon c_-\left(\partial_-\widehat{\mathcal{S}} \times [0,1]\right) \xrightarrow{\cong} C_-(\partial_- H \times [0,1])$$

so that $\Psi|_{\partial_-\widehat{\mathcal{S}}} = \Gamma|_{\partial_-\widehat{\mathcal{S}}}\colon \partial_-\widehat{\mathcal{S}} \xrightarrow{\cong} \partial_- H$. In particular, if $\Phi\colon S^1 \times D^2 \xrightarrow{\cong} \partial_-\widehat{\mathcal{S}}$ is the attaching region, we have that $\Psi \circ \Phi\colon S^1 \times D^2 \to \partial_- H = S^1 \times D^2$ is the identity map.

PROOF. This follows from Lemma 28.12 and the fact that $\Psi|_{c_-\left(\partial_-\widehat{\mathcal{S}}\times[0,1]\right)} = \Psi_\varnothing|_{c_-\left(\partial_-\widehat{\mathcal{S}}\times[0,1]\right)}$. For the condition on the attaching region, see (28.1). $\square$

28.5. From holes and gaps to holes$^+$ and gaps$^+$

As ever, our goal is to show that the skyscraper $\widehat{\mathcal{S}}$ is homeomorphic to the standard handle H, relative to the attaching regions. The strategy will involve shrinking the closures of the connected components of the complement of the design in both the standard handle and the skyscraper. As we saw earlier, these components are called holes and gaps respectively. These are the pieces that are left uncharted by the design, and our philosophy is that since we cannot explore these strange regions, we try to crush them instead and hope that the resulting spaces are homeomorphic. For this to be a reasonable strategy, the holes and gaps would have to be cellular, by Proposition 4.21. However, they are not cellular. This is easiest to see in the standard handle: other than the "central" hole that is homeomorphic to the 4-ball, all other holes are homeomorphic to $S^1 \times D^3$. Therefore for the shrinking strategy to work, we must modify the holes. We do this by adding spanning discs to the holes to convert them into *holes*$^+$, which are cellular. The homeomorphism $\Psi\colon \mathfrak{D} \xrightarrow{\cong} D$ produces the spanning discs in $\mathfrak{D}$ as well, and we attach those to the gaps to transform them into *gaps*$^+$. At first glance that this appears to be a slightly backwards step since we give up some pieces that we have already explored, but it turns out that what we get in return is worth it. Certainly we have to try something different, since we have non-cellular holes and gaps at present.

Let us study the design D in the standard handle $H = D^2 \times D^2$. Recall that we work with the ternary Cantor set in $[\frac{2}{3}, 1]$, or in other words the intersection $\mathfrak{C}_3 \cap [\frac{2}{3}, 1]$. By construction, apart from a collar on $\partial_- H$, the design lies entirely in $D^2 \times S^1 \times [\frac{2}{3}, 1]$. Thus $D \cap (D^2 \times S^1_r)$ is just a slice of this collar on $\partial_- H$ for a radius $r < \frac{2}{3}$. For a radius $r \geq \frac{2}{3}$. there are two possibilities. If r is a Cantor set radius, that is if $r = 1 - \frac{1}{3}\Sigma_{j=1}^{\infty} \frac{2i_j}{3^j}$ for some infinite binary word $i_1 i_2 \dots$, then the entire solid torus $D^2 \times S^1_r$ is contained in the design D, by automatic shrinking. If r is not such a Cantor set radius, it must lie in some middle third interval that was removed in the process of constructing the Cantor set. That is, there is some finite binary word I such that $r \in [r_{I\cdot 1}(0), r_{I\cdot 0}(1)]$. In this case, $D \cap (D^2 \times S^1_r)$ is the image of the map ψ_I. We saw earlier that this image is the complement of a tubular neighbourhood of a mixed ramified Bing-Whitehead link in $D^2 \times S^1_r$; call the closure of this neighbourhood L_I. Note that the holes are $L_I \times [r_{I\cdot 1}(0), r_{I\cdot 0}(1)]$, where I ranges over the set of finite binary words. Let h^j_I denote a connected component of $L_I \times [r_{I\cdot 1}(0), r_{I\cdot 0}(1)]$, for $j = 1, \dots, m_I$. Recall that the mixed ramified Bing-Whitehead link has Whitehead doubles at the last stage. Thus ($\pm$2-framed) longitudes of the solid tori

$$\big\{h^j_I \cap (D^2 \times S^1_{r_I(1/2)})\big\}_{j=1}^{m_I}$$

span immersed discs $\{\Delta^j_I \subseteq D^2 \times S^1_{r_I(1/2)}\}$, as shown in Figure 28.8. Recall that

$$r_I(\tfrac{1}{2}) \in (r_{I\cdot 1}(0), r_{I\cdot 0}(1)).$$

We may choose the collection $\{\Delta^j_I\}_{j,I}$ to be mutually disjoint, and moreover to be such that for $\Delta^j_I \cap h^k_J = \emptyset$ whenever $J \neq I$ or $k \neq j$. For $J \neq I$, this follows since $r_I(\frac{1}{2}) \neq r_J(\frac{1}{2})$, and for $I = J, k \neq j$, this follows by our choice of discs. The interior of the discs Δ^j_I, apart from the sub-discs where Δ^j_I intersects h^j_I (as shown in Figure 28.8), are contained within the design D.

Remark 28.16. For the general skyscrapers of Chapter 24, the corresponding immersed discs may not be quite as well behaved. Nonetheless, a similar argument applies.

As mentioned above, we would like to add embedded spanning discs to the holes to convert them into cellular sets. We will find these embedded discs by modifying the immersed discs $\Delta^j_I \subseteq D^2 \times S^1_{r_I(1/2)}$. The idea is that we will push the singularities of the immersed discs in a solid torus at a non-Cantor set radius to a nearby solid torus at a Cantor set radius. We will need to be careful that the new discs do not intersect each other or the holes, which we shall arrange by ensuring that no two discs use the same Cantor set radius, and by controlling how we push. Since the

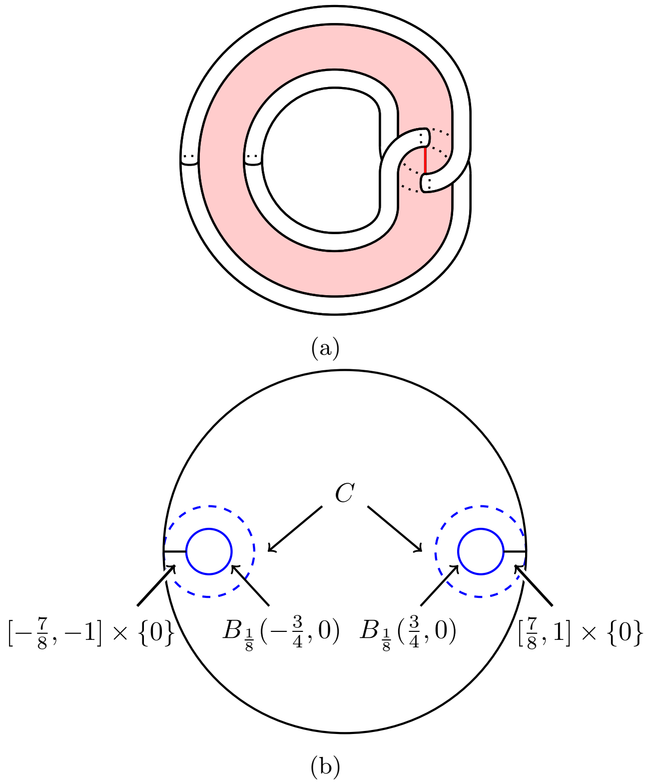


Figure 28.8. (a) A spanning disc of a Whitehead double. (b) The intersections shown within the domain D^2, including the parametrisation of the intersections.

Cantor set is uncountable and the number of holes is countable, we do have some room for manoeuvre.

The disc Δ_I^j intersects h_I^j in two disjoint sub-discs, and itself in two intervals connecting these sub-discs with the boundary. Parametrise the disc $\Delta_I^j \cong D^2$ as the unit disc in $\mathbb{R}^2$. Choose the parametrisation of Δ_I^j so that it intersects h_I^j in the sub-discs $B_{\frac{1}{8}}(\pm 3/4, 0)$ and itself in the intervals $[-1, -\frac{7}{8}] \times \{0\}$ and $[\frac{7}{8}, 1] \times \{0\}$. Let

$$C := \mathring{B}_{\frac{1}{4}}(-3/4, 0) \cup \mathring{B}_{\frac{1}{4}}(3/4, 0)$$

and let $b\colon D^2 \to [0,1]$ be a smooth function with $b \equiv 1$ on $B_{\frac{1}{8}}(\pm 3/4, 0)$, $b > 0$ on C, and $b \equiv 0$ on $D^2 \setminus C$; see Figure 28.8(b) and Figure 28.9. We give a recursive algorithm to perturb the immersed discs Δ_I^j to a collection of mutually disjoint embedded discs within the design D. To begin, let $\Upsilon := \{\frac{2}{3}, 1\}$. This set gives the Cantor set radii we must avoid at each step. Thus whenever we use a Cantor set radius, we shall add it to this set. In other words, the set Υ will be updated as the algorithm progresses.

At each step, choose a finite binary word I of shortest length for which we do not yet have spanning discs for the corresponding holes. Recall the immersed discs $\{\Delta_I^j\}_j$ corresponding to a finite binary word I lie in $D^2 \times S^1_{r_I(1/2)}$ where $r_{I\cdot 1}(0) < r_I(\frac{1}{2}) < r_{I\cdot 0}(1)$. Choose points $d_I, \widehat{d}_I$ in the Cantor set such that

$$r_{I\cdot 0}(1) < d_I < \min\{\Upsilon \cap [r_{I\cdot 0}(1), 1]\},$$

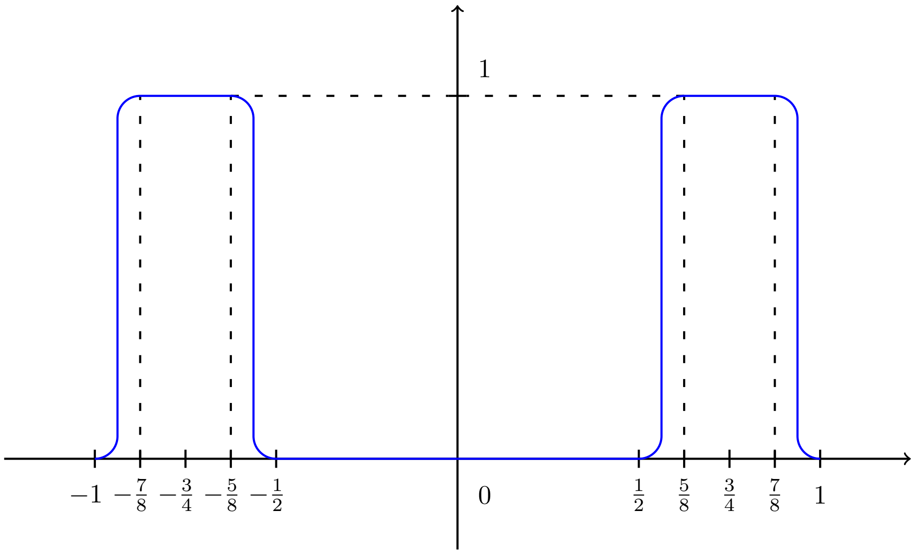


Figure 28.9. A graph of the function b restricted to the abscissa.

$$r_{I\cdot 1}(0) > \widehat{d}_I > \max\left\{\Upsilon \cap \left[\frac{2}{3}, r_{I\cdot 1}(0)\right]\right\},$$

and such that d_I and $\widehat{d}_I$ are not endpoints of any middle third interval. Define $\phi_I\colon D^2 \to [0,1]$ by

$$\phi_I(s) = \begin{cases} r_I(\frac{1}{2}) + b(s)(d_I - r_I(\frac{1}{2})) & \text{if } s \in \mathring{B}_{\frac{1}{4}}(3/4, 0), \\ r_I(\frac{1}{2}) + b(s)(\widehat{d}_I - r_I(\frac{1}{2})) & \text{if } s \in \mathring{B}_{\frac{1}{4}}(-3/4, 0), \\ r_I(\frac{1}{2}) & \text{otherwise.} \end{cases}$$

Note that $\widehat{d}_I < r_I(\frac{1}{2}) < d_I$, so part of the disc is perturbed in each direction. For each element of $\{h_I^j\}_j$ corresponding to I, take the parametrised discs $\Delta_I^j\colon D^2 \to D^2 \times S^1$ as above and embed discs by

$$\widetilde{\Delta}_I^j\colon D^2 \xrightarrow{\Delta_I^j \times \phi_I} D^2 \times S^1 \times [\widehat{d}_I, d_I]$$

Add the points $\{d_I, \widehat{d}_I, r_{I\cdot 1}(0), r_{I\cdot 0}(1)\}$ to Υ.

Performing this recursion gives embedded spanning discs for all the components of all the holes, since they are indexed by finite binary words. The lemma below shows that the process described above is possible and effective.

Lemma 28.17. *The radii d_I and $\widehat{d}_I$ may be chosen as described above for all finite binary words I. The resulting discs $\widetilde{\Delta}_I^j$ are mutually disjoint and embedded within the design D.*

Proof. In the algorithm described above, for each finite binary word I, we seek $d_I \in \mathfrak{C}_3 \cap [\frac{2}{3}, 1]$ such that $r_{I\cdot 0}(1) < d_I < \min\{\Upsilon \cap [r_{I\cdot 0}(1), 1]\}$, and d_I is not an endpoint of any middle third interval. Here Υ is a recursively defined finite set consisting of elements of the Cantor set as defined earlier. The point $r_{I\cdot 0}(1)$ does not lie in Υ since it is a boundary point of a middle third interval, and thus cannot be $d_{I'}$, nor $\widehat{d}_{I'}$, for some finite binary word I'. Nor can it be written in the form $r_{I''\cdot 1}(0)$ or $r_{I''\cdot 0}(1)$ for some finite binary word $I'' \neq I$. Thus we are at least not seeking an element of the empty set. The set Υ is finite at each stage and so the minimal element is defined.

Note that $r_{I\cdot 0}(0) \in \Upsilon$. This follows since $I \cdot 0$ is either $00\ldots 0$ or can be written as $\bar{I} \cdot 1 \cdot 0 \ldots 0$ for some finite binary word $\bar{I}$. In the first case, $r_{I\cdot 0}(0) = 1 \in \Upsilon$. In the latter case, $r_{I\cdot 0}(0) = r_{\bar{I}\cdot 1}(0)$, and since $\bar{I}$ has smaller length than I, $r_{I\cdot 0}(0) = r_{\bar{I}\cdot 1}(0) \in \Upsilon$. Thus we seek a $d_I \in (r_{I\cdot 0}(1), r_{I\cdot 0}(0))$.

If Υ has no intersection with $(r_{I\cdot 0}(1), r_{I\cdot 0}(0))$, we easily choose d_I satisfying the given conditions. Suppose that Υ has nontrivial intersection with $(r_{I\cdot 0}(1), r_{I\cdot 0}(0))$. Let d be the minimal intersection point. From the definition of the set Υ we see that d must be d_J or $\widehat{d}_J$ for some J, in fact some element of the Cantor set which is not the endpoint of a middle third interval. The last condition requires some thought to verify and uses the fact that we are working with a finite binary word of minimal length which does not yet have a spanning disc. We must choose an element $d_I \in \mathfrak{C}_3 \cap (r_{I\cdot 0}(1), d)$. The next paragraph describes an explicit method to do so.

We know that such a d is uniquely represented, via the map $\mathfrak{c}$ from Section 28.1, as an infinite binary word that is not essentially finite. Since $r_{I\cdot 0}(1)$ is the left boundary point of a middle third interval, it can be represented by an infinite binary word of the form $N \cdot 00\ldots$, for some N with minimal length. Since $d \in (r_{I\cdot 0}(1), r_{I\cdot 0}(0))$, the word for d starts with N, that is it is given by $N \cdot N'$ for some essentially infinite (i.e. non-essentially finite) binary word N'. Note that N' must have an infinite number of 1s since it is essentially infinite. Let $\widetilde{N'}$ be the result of changing any one of those 1s to a 0. Then $N \cdot \widetilde{N'}$ is also an infinite binary word which is not essentially finite, and thus gives an element of the Cantor set that is not a boundary of a middle third interval. By construction, this element lies between $r_{I\cdot 0}(1)$ and d, and we call it d_I.

A similar argument shows that we can find $\widehat{d}_I$ for any finite binary word I and that $\widehat{d}_I \in (r_{I\cdot 1}(1), r_{I\cdot 1}(0))$. It is easy to see by construction that if d_K or $\widehat{d}_K$ lies in $(\widehat{d}_I, d_I)$, for some finite binary words I and K, then $|I| < |K|$ and moreover, $[\widehat{d}_K, d_K] \subseteq (\widehat{d}_I, d_I)$.

We need to check that the discs $\{\widetilde{\Delta}_I^j\}$ are mutually disjoint and embedded within the design. Recall that the original immersed discs Δ_I^j were mutually disjoint (although they had self-intersections) and located within the design, apart from the sub-discs where each $\widetilde{\Delta}_I^j$ intersected the hole h_I^j containing $\partial\Delta_I^j$, for finite I, and j. Thus we only need to check what happens when we apply the functions ϕ_I. Most of the disc Δ_I^j is unaltered: only the points in C are pushed towards the nearby Cantor set radii d_I and $\widehat{d}_I$. The image of the points in C lies in the set $D^2 \times S^1 \times [\widehat{d}_I, d_I]$. Then discs $\widetilde{\Delta}_I^j$ and $\widetilde{\Delta}_I^k$, for $j \neq k$ and fixed I are disjoint by the construction of the function ϕ_I, since the projections of the discs to $D^2 \times S^1 \times \{r_I(\frac{1}{2})\}$ are disjoint.

By construction, the discs Δ_I^j and Δ_K^ℓ are disjoint as long as $[\widehat{d}_I, d_I]$ and $[\widehat{d}_K, d_K]$ are disjoint. It remains only to convince ourselves that if there exists K such that d_K or $\widehat{d}_K$ lies in $(\widehat{d}_I, d_I)$, the discs $\widetilde{\Delta}_I^i$ and $\widetilde{\Delta}_K^j$ do not intersect. This is implied by the compatibility condition (H3) on the ψ maps. To see this, since d_K or $\widehat{d}_K$ lies in $(\widehat{d}_I, d_I)$, we have that $|I| < |K|$ and $[\widehat{d}_K, d_K] \subseteq (\widehat{d}_i, d_I)$. Then $\{\widetilde{\Delta}_K^\ell\}$ lies in $L_K \times [\widehat{d}_K, d_K] \subseteq D^2 \times S^1 \times [\widehat{d}_K, d_K]$ while $\{\widetilde{\Delta}_I^j\}$ lies in $L_I \times [\widehat{d}_I, d_I] \subseteq D^2 \times S^1 \times [\widehat{d}_I, d_I]$. By the compatibility condition, L_K is contained strictly within L_I, which shows that the collections of discs are mutually disjoint.

Finally, we check that the discs $\{\widetilde{\Delta}_I^j\}_{I,j}$ lie within the design. This is straightforward by construction. That is, away from the radii $\widehat{d}_I$ and d_I, the image of C is embedded within the design. But this also holds within the Cantor set radii, since by automatic shrinking the entire solid torus at a Cantor set radius in $[\frac{2}{3}, 1]$ is contained in the design. $\square$

We will need the following in the proof of the α shrink in Proposition 28.21.

SCHOLIUM 28.18. *Let I be a finite binary word. Then $[\widehat{d}_I, d_I] \subseteq (r_I(1), r_I(0))$.*

PROOF. We saw in the proof of Lemma 28.17 that $d_I \in (r_{I\cdot 0}(1), r_{I\cdot 0}(0))$ and $\widehat{d}_I \in (r_{I\cdot 1}(1), r_{I\cdot 1}(0))$. Note that $r_{I\cdot 1}(1) = r_I(1)$ and $r_{I\cdot 0}(0) = r_I(0)$. $\square$

We will need the following in the proof of the β shrink in Section 28.6.3.

LEMMA 28.19. *Let I be a finite binary word. If a hole$^+$ corresponding to a finite binary word J intersects the region $D^2 \times S^1 \times [r_{I\cdot 1}(0), r_{I\cdot 0}(1)]$, then $|J| \leq |I|$.*

PROOF. Observe that $(r_{I\cdot 1}(0), r_{I\cdot 0}(1))$ is one of the middle third intervals removed in the construction of $\mathfrak{C}_3 \cap [\frac{2}{3}, 1]$. So for any J with $|J| > |I|$, we have either that $r_J(\frac{1}{2}) > r_{I\cdot 0}(1)$ or $r_J(\frac{1}{2}) < r_{I\cdot 1}(0)$. Recall also that the points $r_{I\cdot 1}(0)$ and $r_{I\cdot 0}(1)$ were added to the set Υ in the step where we perturbed the spanning discs for the holes corresponding to the word I. Thus the perturbations of the spanning discs for holes corresponding to J cannot enter the region $D^2 \times S^1 \times [r_{I1}(0), r_{I0}(1)]$. $\square$

Define
$$h_I^{j\,+} := h_I^j \cup \widetilde{\Delta}_I^j \subseteq D^2 \times D^2,$$
for each I and j, and call the resulting subsets, along with the central hole, the *holes*$^+$. The collection of holes$^+$ determines a decomposition of the standard handle H, denoted $\mathcal{H}^+$. Recall that the discs $\widetilde{\Delta}_I^j$ are contained in the design D. They are the *red blood cell discs*. Use the homeomorphism $\Psi\colon \mathfrak{D} \xrightarrow{\cong} D$ to map them to the skyscraper $\widehat{\mathcal{S}}$. The union of the gaps with the images of the red blood cell discs in $\widehat{\mathcal{S}}$ are called *gaps*$^+$; the collection of gaps$^+$ forms a decomposition of $\widehat{\mathcal{S}}$, denoted $\mathcal{G}^+$.

Finally, by definition, the boundary ∂h_I^j of each hole h_I^j, is contained within the design $\mathfrak{D}$, as are the red blood cell discs $\{\widetilde{\Delta}_I^j\}_{I,j}$. The connected components of the union of these two collections forms a decomposition, denoted $\mathcal{D}$, of the design $\mathfrak{D}$.

28.6. Shrinking the complement of the design

In this last step of the proof, we focus on the following three decompositions.

- The collection of holes$^+$, denoted $\mathcal{H}^+$, in the standard handle H.
- The collection of gaps$^+$, denoted $\mathcal{G}^+$, in the skyscraper $\widehat{\mathcal{S}}$.
- The collection of connected components of boundaries of holes union red blood cell discs, denoted $\mathcal{D}$, in the design $\mathfrak{D}$.

In the third decomposition, we are considering the design $\mathfrak{D}$ as an abstract object, rather than being embedded in any ambient space.

28.6.1. The common quotient. The three decompositions above have a "common quotient" in the following sense.

LEMMA 28.20. *The quotients $H/\mathcal{H}^+$ and $\widehat{\mathcal{S}}/\mathcal{G}^+$ are each canonically homeomorphic to $\mathfrak{D}/\mathcal{D}$.*

PROOF. By construction the collection of holes/gaps is in bijective correspondence with the collection of holes$^+$/gaps$^+$. The collections of red blood cell discs in the abstract design, as well as in the design in the skyscraper and the design in the standard handle are explicitly identified. So it will suffice to show that the collection of holes and the collection of gaps are each in bijective correspondence with the collection of boundary components of the holes, forming part of the decomposition $\mathcal{D}$, within the abstract design $\mathfrak{D}$. This is obviously true for the holes from the explicit parametrisation.

In addition, each gap meets at least one boundary component as follows. Observe that the design in the skyscraper is closed, so the complement in the skyscraper is a disjoint union of open sets such that the closure of each is a gap. But a gap and the design have to meet along the topological boundary (meaning limit points minus interior points) of the design, unless one of those complementary pieces were closed to begin with. This would mean that it was both closed and open, which contradicts that the skyscraper is connected. Therefore each gap meets at least one boundary component. But *a priori*, it is possible that a single gap meets multiple boundary components.

We shall show that each gap meets at most one boundary component which will complete the proof. Note that there is a central gap corresponding to the boundary of the central hole in the standard handle. No other gap meets that boundary component (the vertical boundary $\partial_+\widehat{\mathcal{S}}_{111\ldots} \subseteq \mathfrak{D}$ separates it from all other gaps (Lemma 28.11)), so we ignore it in the rest of the proof. Before addressing the other gaps, we gather some relevant facts. Consider a boundary component ∂h_I^j of the design for some finite binary word I. We use this notation, coming from the notation for holes in the previous section, since the boundary of a hole is within the design. In other words, we use the same notation here for the image of these boundaries under the homeomorphism $\Psi^{-1}\colon D \xrightarrow{\cong} \mathfrak{D}$.

(Q1) For binary words I and J, $\mathfrak{c}_I < \mathfrak{c}_J$ if and only if $\widehat{\mathcal{S}}_J \subsetneq \widehat{\mathcal{S}}_I$, in which case $\partial_+\widehat{\mathcal{S}}_J$ separates $\widehat{\mathcal{S}}_I$ (Lemma 28.11).

(Q2) For each I and j, ∂h_I^j and any gap meeting it lie in $\widehat{\mathcal{S}}_{I\cdot 0111\ldots} \smallsetminus \operatorname{Int}\widehat{\mathcal{S}}_{I\cdot 1}$, above the first $|I|+1$ levels of $\widehat{\mathcal{S}}_{I\cdot 0111\ldots}$. These are obtained from the first $|I|+1$ levels of $\widehat{\mathcal{S}}_I$ by removing part of the collar c_I of $\partial_+\widehat{\mathcal{S}}_I$ (Remark 28.9). We are using the fact that $\partial_+\widehat{\mathcal{S}}_{I\cdot 0111\ldots}$ and $\partial_+\widehat{\mathcal{S}}_{I\cdot 1}$ are both contained in $\mathfrak{D}$ and separate $\widehat{\mathcal{S}}_I$ (by (Q1)), as well as that $\widehat{\mathcal{S}}_I^{\leq|I|+1} \cap (\widehat{\mathcal{S}}_{I\cdot 0111\ldots}^{\leq|I|+1} \smallsetminus \operatorname{Int}\widehat{\mathcal{S}}_{I\cdot 1}^{\leq|I|+1}) \subseteq \mathfrak{D}$. More precisely, the boundary ∂h_I^j is contained in the union
$$\big(\partial_+\widehat{\mathcal{S}}_{I\cdot 0111\ldots}\big) \cup \big(\partial_+\widehat{\mathcal{S}}_{I\cdot 1}\big) \cup \big(\widehat{\mathcal{S}}_I^{\leq|I|+1} \cap (\widehat{\mathcal{S}}_{I\cdot 0111\ldots}^{\leq|I|+1} \smallsetminus \operatorname{Int}\widehat{\mathcal{S}}_{I\cdot 1}^{\leq|I|+1})\big),$$
and the latter set being contained in the design means that the gap has to lie above the first $|I|+1$ levels of $\widehat{\mathcal{S}}_{I\cdot 0111\ldots}$ as asserted. Note that ∂h_I^j intersects the tip region of $\widehat{\mathcal{S}}_{I\cdot 0111\ldots}^{|I|+1}$.

We begin the argument. Consider two boundary components ∂h_I^j and ∂h_J^k, for the finite binary words $I \neq J$. By Lemma 28.1, either $\mathfrak{c}_{I\cdot 1} < \mathfrak{c}_{J\cdot 0111\ldots}$ or $\mathfrak{c}_{J\cdot 1} < \mathfrak{c}_{I\cdot 0111\ldots}$. First assume that $\mathfrak{c}_{I\cdot 1} < \mathfrak{c}_{J\cdot 0111\ldots}$. Then by Lemma 28.2, there is some $c \in \mathfrak{C}_3$ with $\mathfrak{c}_{I\cdot 1} < c < \mathfrak{c}_{J\cdot 0111\ldots}$. Let L be the (unique) infinite binary word associated with c. Then $\widehat{\mathcal{S}}_{J\cdot 0111\ldots} \subsetneq \widehat{\mathcal{S}}_L \subsetneq \widehat{\mathcal{S}}_{I\cdot 1} \subsetneq \widehat{\mathcal{S}}_I$ by (Q1). In particular, $\partial_+\widehat{\mathcal{S}}_L$ lies in the design in the skyscraper and separates $\widehat{\mathcal{S}}_I$ such that ∂h_I^j and ∂h_J^k lie in different components. A virtually identical argument applies when $I \neq J$ and $\mathfrak{c}_{J\cdot 1} < \mathfrak{c}_{I\cdot 0111\ldots}$. In both cases, a gap can meet at most one of the two boundary components.

It remains to consider the case of boundary components ∂h_I^j and ∂h_I^k for $j \neq k$ and fixed I. By (Q2), the boundary component ∂h_I^j and any gap that meets it are contained in $\widehat{\mathcal{S}}_{I\cdot 0111\ldots}$ above the first $|I|+1$ levels. Remove the first $|I|+1$ levels from $\mathcal{S}_{I\cdot 0111\ldots}$. We are left with a mutually disjoint union of skyscrapers, the collection of whose attaching regions within the boundary of the first $|I|+1$ levels is in one to one correspondence with the set $\{\partial h_I^j\}_j$. This last statement follows once again by comparing the construction of the design in the skyscraper and the design in the standard handle. In other words, each skyscraper within this mutually disjoint collection contains precisely one boundary component of a hole. Any gap meeting such a boundary component must therefore lie in the same skyscraper. This completes the proof that no gap can meet two distinct boundary components. □

Given Lemma 28.20, we know that the collection of holes and the collection of gaps are in bijective correspondence, defined using the boundary components and the homeomorphism $\Psi\colon \mathfrak{D} \xrightarrow{\cong} \mathcal{D}$. Let g_I^j denote the gap corresponding to h_I^j, so that $\Psi(\partial g_I^j) = \partial h_I^j$.

Let Q denote the quotient $\mathfrak{D}/\mathcal{D}$ and consider the following commutative diagram:

$$\begin{tikzcd}
H \arrow[d, "/\mathcal{H}^+"'] \arrow[drr, "\alpha"] & D \arrow[l, hook'] & \mathfrak{D} \arrow[l, "\Psi"', "\cong"] \arrow[d, "/\mathcal{D}"] \arrow[rr, hook] & & \widehat{\mathcal{S}} \arrow[d, "/\mathcal{G}^+"] \arrow[dll, "\beta"'] \\
H/\mathcal{H}^+ \arrow[rr, "\cong"] & & \mathfrak{D}/\mathcal{D} =: Q & & \widehat{\mathcal{S}}/\mathcal{G}^+ \arrow[ll, "\cong"']
\end{tikzcd} \tag{28.8}$$

where α and β denote the compositions of the lower triangles, and the horizontal homeomorphisms in the bottom row are from the previous lemma. The vertical maps are the quotient maps. We will finish the proof by showing that α and β are approximable by homeomorphisms, using our knowledge of decomposition space theory.

28.6.2. The map α is approximable by homeomorphisms.

Proposition 28.21 (The α shrink). *The decomposition $\mathcal{H}^+$ of $H = D^2 \times D^2$ given by the holes$^+$ is shrinkable, relative to the boundary union the collar $C_-(\partial_- H \times [0, \frac{1}{2}])$.*

In other words, the map $\alpha\colon H \to Q$ is approximable by homeomorphisms, agreeing with α on $\partial H \cup C_-(\partial_- H \times [0, \frac{1}{2}])$.

Proof. This will follow from the starlike null theorem (Theorem 26.4), which states that null, recursively starlike-equivalent decompositions of H shrink. By construction the elements of the decomposition $\mathcal{H}^+$ are located away from $\partial H \cup C_-(\partial_- H \times [0, \frac{1}{2}])$. Recall that each hole$^+$ is recursively starlike-equivalent, with one recursion by Lemma 26.3. So we only need to show that $\mathcal{H}^+$ is null. Since null decompositions are upper semi-continuous (Lemma 9.3), this also shows that the quotient Q is metrisable (Corollary 4.13). Recall that a collection of subsets of a metric space is said to be *null* if for any given $\varepsilon > 0$ there are only finitely many subsets with diameter greater than ε (Definition 9.1). Since the diameter of a hole$^+$ is at most the sum of the diameter of the corresponding hole and spanning disc, it suffices to show that the collection of holes and the collection of spanning discs are individually null.

First we address the collection of holes, arguing by contradiction. Suppose that there is an $\varepsilon > 0$ such that infinitely many holes have diameter at least ε. For ease of reference, we say that a set is *large* if its diameter is at least ε. As in Section 28.5, we denote the collection of holes by $\{h_I^j\}_{I,j}$, where there are finitely many h_I^j for each finite binary word I. We saw in that section that each h_I^j is a connected component of $L_I \times [r_{I\cdot 1}(0), r_{I\cdot 0}(1)]$, where L_I in $D^2 \times S^1$ is a thickened mixed ramified Bing-Whitehead link. Let $\{L_I^j\}_j$ be the set of components of L_I. The length of the interval $[r_{I\cdot 1}(0), r_{I\cdot 0}(1)]$ is $\frac{1}{3^{|I|+2}}$ and thus the collection of intervals is null in the interval $[0, 1]$. If there are infinitely many large holes, there must be infinitely many large solid tori L_I^j. Moreover, according to (H3) the links $\{L_I\}_I$ are nested in the sense that if $I \preceq J$, then $L_J \subseteq L_I$, so that each component of L_J is contained in some component of L_I and in particular, if L_J has a large component, then so does L_I. Call a solid torus *huge* if it contains infinitely many large solid tori.

To obtain a contradiction we will construct an infinite binary word $i_1 i_2 i_3 \cdots \in \{0,1\}^\infty$ that violates the automatic shrinking property of Lemma 28.14. We construct this word inductively along with a sequence of huge solid tori $L_k = L_{i_1 i_2 \ldots i_k}^{j_k}$ such that $L_{k+1} \subsetneq L_k$. Since there are infinitely many large solid tori, there must be some choice of j_1 and a choice of $i_1 \in \{0,1\}$ such that $L_{i_1}^{j_1}$ is huge (recall that there are only finitely many solid tori corresponding to binary words with length one). Call this solid torus L_1. Now assume that we have already chosen $i_1, \ldots, i_k$ and $L_1, \ldots, L_k$ for some $k \geq 1$. Then one of the finitely many solid tori of the form $L_{i_1 i_2 \ldots i_k \ell}^j \subseteq L_k$ must be huge. We denote this solid torus by L_{k+1} and set i_{k+1} to be ℓ. Since each of the L_k are huge, the intersection $\bigcap_k L_k$ must be large. However, by automatic shrinking, $\bigcap_k L_k$ must be a point since it is one of the decomposition elements in Lemma 28.14. This gives the required contradiction.

It remains to address the collection of spanning discs. Let J denote I with the last digit deleted. Observe that the diameter of each unperturbed disc, namely the immersed spanning disc Δ_I^j for a Whitehead curve, is bounded by the diameter of a (tubular neighbourhood of a) link component $L_J^{j'}$, with j' such that $L_I^j \subseteq L_J^{j'}$. Therefore the unperturbed spanning disc has diameter bounded by the diameter of the hole $h_J^{j'}$. We already showed that the collection of holes is null. Therefore we only need to verify that the perturbations by which we made the spanning discs embedded do not increase the size of the discs too much. Recall that the perturbations were constructed inductively and were contained within a subinterval of $[0,1]$. These subintervals are contained in middle third intervals used in the Cantor set construction of strictly decreasing length (Scholium 28.18). Consequently the collection of subintervals where the perturbations take place are null. It follows that the resulting collection of embedded spanning discs is also null. $\square$

28.6.3. The map β is approximable by homeomorphisms. The next lemma and the following proposition are straightforward point set topology results that we will use in the proof of the β shrink, which will complete the proof of the disc embedding theorem.

LEMMA 28.22. *Let d_1, d_2 be two metrics on a compact space X that induce the same topology. Then a collection $\mathcal{F}$ of subsets of X is null with respect to d_1 if and only if it is null with respect to d_2.*

PROOF. Since d_1 and d_2 induce the same topology on X, the identity map $\mathrm{Id}\colon (X, d_1) \to (X, d_2)$ a continuous map of metric spaces. Since X is compact, by the Heine-Cantor theorem (Theorem 3.25) the identity map is uniformly continuous.

Now assume that the decomposition $\mathcal{F}$ is null with respect to d_1. We show that $\mathcal{F}$ is also null with respect to d_2. By symmetry, this prove the lemma. Fix $\varepsilon > 0$. Since Id is uniformly continuous, there exists $\delta > 0$ so that for every $A \subseteq X$, $\operatorname{diam}_{d_2} A < \varepsilon$ whenever $\operatorname{diam}_{d_1} A < \delta$. Since $\mathcal{F}$ is null with respect to d_1, there are only finitely many elements $F_1, \ldots, F_n$ of $\mathcal{F}$ with diameter greater than δ (with respect to d_1). Then all elements of $\mathcal{F} \setminus \{F_1, \ldots, F_n\}$ have diameter less than ε with respect to d_2. This completes the proof. $\square$

PROPOSITION 28.23. *Let $\mathcal{F}$ be a decomposition of a space X such that for every element $D \in \mathcal{F}$, and every open neighbourhood U of D, there is a nonempty open* *subset $V \subseteq U$ that does not meet any element of $\mathcal{F}$ (including D). Then the image of the (non-singleton) decomposition elements in $X/\mathcal{F}$ is nowhere dense.*

Proof. Without loss of generality, assume that each element of $\mathcal{F}$ is a non-singleton set. Recall that for $A \subseteq X$ the closure $\mathrm{Cl}(A)$ is the set of limit points of A, where a point $x \in X$ is a limit point of A if every open set containing x also contains a point of A (which might be x itself). A subset of a topological space is said to be nowhere dense if its closure has empty interior. Let $f\colon X \to X/\mathcal{F}$ be the quotient map. We will show that the closure of the interior of $A := \bigcup_{D\in\mathcal{F}} f(D)$, the set of singular points in $X/\mathcal{F}$, is empty.

Let $x \in \mathrm{Cl}(A)$ and let W be an open neighbourhood of x in $X/\mathcal{F}$. We will show that W is not contained in $\mathrm{Cl}(A)$, which will complete the proof. Since x is a limit point of A, the open set W contains some element of A. That is, there is a $D \in \mathcal{F}$ such that $f(D) \in W$. Thus $f^{-1}(W)$ is an open neighbourhood of D. By hypothesis there is a nonempty open set $V \subseteq f^{-1}(W)$ such that $V \cap D' = \emptyset$ for all $D' \in \mathcal{F}$.

Now let $y \in V$. The restriction $f|_V$ is a homeomorphism onto its image since V does not meet any decomposition element. Hence $f(V)$ is an open neighbourhood of $f(y) \in f(V) \subseteq W$ and $f(V)$ contains no singular point $f(D')$. It follows that $f(y)$ is not a limit point of $A = \bigcup_{D'\in\mathcal{F}} f(D')$. Since W contains $f(y)$, this implies that W is not contained in $\mathrm{Cl}(A)$, as needed. □

By Proposition 28.21 we know that the common quotient Q is homeomorphic to the 4-ball, and we wish to apply the ball to ball theorem (Theorem 26.5). We meet an immediate obstacle since we do not yet know that $\widehat{\mathcal{S}}$ is a 4-ball. To circumvent this we use the collar adding lemma (Lemma 26.11).

Lemma 26.11 (Collar adding lemma). *The union $\widehat{\mathcal{S}} \cup_{\mathrm{Id}} (\partial\widehat{\mathcal{S}} \times [0,1])$ is homeomorphic to $D^2 \times D^2$.*

We remark that we are not claiming the existence of such a collar in an ambient 4-manifold. Indeed, if such a collar could always be found *a priori* in the ambient space, there would be a significantly simpler proof of the disc embedding theorem.

We gather some relevant facts for the rest of the proof.

(B1) The map $\alpha\colon H \to Q$ is a homeomorphism when restricted to ∂H union the collar $C_-(\partial_- H \times [0,\frac{1}{2}])$. This follows from the definition of α and the fact that no elements of $\mathcal{H}^+$ meet this region.

(B2) The map $\beta\colon \widehat{\mathcal{S}} \to Q$ is a homeomorphism restricted to $\partial\widehat{\mathcal{S}}$ union the collar $c_-(\partial_-\widehat{\mathcal{S}} \times [0,\frac{1}{2}])$. This follows from the definition of β and the fact that no elements of $\mathcal{G}^+$ meet this region.

(B3) The map
$$\alpha^{-1}\circ\beta|_{c_-\left(\partial_-\widehat{\mathcal{S}}\times\left[0,\frac{1}{2}\right]\right)}\colon c_-(\partial_-\widehat{\mathcal{S}} \times [0,\tfrac{1}{2}]) \to C_-(\partial_- H \times [0,\tfrac{1}{2}])$$
agrees with $\Psi\colon \mathfrak{D} \xrightarrow{\cong} D$ and is in particular a diffeomorphism. See (28.8) and Lemma 28.15.

(B4) The restriction $\alpha^{-1}\circ\beta|_{\partial_-\widehat{\mathcal{S}}}$ agrees with the restriction $\Psi|_{\partial_-\widehat{\mathcal{S}}}\colon \partial_-\widehat{\mathcal{S}} \xrightarrow{\cong} \partial_- H$, so that if $\Phi\colon S^1 \times D^2 \xrightarrow{\cong} \partial_-\widehat{\mathcal{S}}$ is the attaching region, the composition
$$\alpha^{-1}\circ\beta\circ\Phi\colon S^1 \times D^2 \to \partial_- H = S^1 \times D^2$$
is the identity map. See (28.8) and Lemma 28.15.

Glue
$$\mathcal{E} := \partial\widehat{\mathcal{S}} \times [0,1]$$

to $\widehat{\mathcal{S}}$ by identifying each point $(x,0) \in \partial\widehat{\mathcal{S}} \times [0,1]$ to the point $x \in \partial\widehat{\mathcal{S}}$. The resulting space is of course $\widehat{\mathcal{S}} \cup_{\mathrm{Id}} (\partial\widehat{\mathcal{S}} \times [0,1])$, which we know from Lemma 26.11 to be homeomorphic to D^4. Let

$$\daleth \colon D^4 \to \widehat{\mathcal{S}} \cup_{\mathrm{Id}} (\partial\widehat{\mathcal{S}} \times [0,1])$$

be the homeomorphism provided by the collar adding lemma (Lemma 26.11). Use [**Hau30**] to extend the fixed metric on $\widehat{\mathcal{S}}$ to the larger space $\widehat{\mathcal{S}} \cup_{\mathrm{Id}} (\partial\widehat{\mathcal{S}} \times [0,1])$. Via the inclusion map, the decomposition $\mathcal{G}^+$ of $\widehat{\mathcal{S}}$ induces a decomposition of $\widehat{\mathcal{S}} \cup_{\mathrm{Id}} \mathcal{E}$. By a mild abuse of notation, we use the symbol $\mathcal{G}^+$ for this decomposition as well.

LEMMA 28.24. *The decomposition $\mathcal{G}^+$ of $\widehat{\mathcal{S}} \cup_{\mathrm{Id}} \mathcal{E}$ is null.*

PROOF. Since each element of $\mathcal{G}^+$ is contained in $\widehat{\mathcal{S}}$ and by definition of the metric on $\widehat{\mathcal{S}} \cup_{\mathrm{Id}} \mathcal{E}$, it will suffice to prove that the original $\mathcal{G}^+$ is a null decomposition of $\widehat{\mathcal{S}}$. Henceforth in this proof we work in $\widehat{\mathcal{S}}$.

The central gap$^+$ does not affect nullity so we ignore it. Moreover, the red blood cell discs in $\widehat{\mathcal{S}}$ were created by pushing the spanning discs within the design. We may therefore consider the preimage of the red blood cell discs in the standard handle H under the homeomorphism $D \cong \mathfrak{D}$. We already showed that the red blood cell discs in $D \subsetneq H$ are null in the proof of Lemma 28.21. It follows that the red blood cell discs are null in $\mathfrak{D} \subsetneq \widehat{\mathcal{S}}$, with respect to the metric pulled back from $D \subsetneq H$ using the homeomorphism $\mathfrak{D} \cong D$. Since the design is compact, we may check nullity using any metric that induces the same topology, by Lemma 28.22. So the red blood cell discs are also null with respect to the metric we originally fixed on $\widehat{\mathcal{S}}$ (restricted to $\mathfrak{D}$). Just as in the proof of Proposition 28.21, it therefore suffices to show that the gaps are null.

We observed in the proof of Lemma 28.20 that the gap g_I^j corresponding to the hole h_I^j, for some fixed I and j, is contained in $\widehat{\mathcal{S}}_{I\cdot 0111\ldots} \subseteq \widehat{\mathcal{S}}_I$ above the first $|I|+1$ levels (which coincide modulo boundary collars). The proof of nullity will therefore be complete once we show that the set of closures of connected components of complements of certain finite truncations of skyscrapers corresponding to finite binary words is null. More precisely, we wish to show that the following set is null:

$$\mathcal{C} := \big\{\overline{C} \mid C \text{ connected component of } \widehat{\mathcal{S}}_I \smallsetminus \widehat{\mathcal{S}}_I^{\leq |I|+1},\ |I| < \infty\big\}.$$

Fix $\varepsilon > 0$. Let N be the smallest integer such that $\frac{1}{\varepsilon} \leq N$. For each finite binary word I with a 1 as the kth digit for $N < k \leq |I|$, the connected components of the $(k+2)$th and higher levels of $\widehat{\mathcal{S}}_I$ are located in mutually disjoint balls of diameter at most $1/k$ according to Scholium 28.5. In particular, for a finite binary word with a 1 in the kth digit with $N < k \leq |I|$, the corresponding elements of the set $\mathcal{C}$, and therefore the corresponding gaps, have diameter strictly less than ε, since $\frac{1}{k} < \frac{1}{N} \leq \varepsilon$.

It remains only to address the elements of $\mathcal{C}$ corresponding to finite binary words with no 1s beyond the Nth digit. Recall from Section 28.3 that adding a 0 to a finite binary word does not change the corresponding skyscraper. So it remains only to consider the collection of 2^N skyscrapers corresponding to the finite binary words of length N. That is, we need to show that the following collection is null:

$$\mathcal{C}_N := \big\{\overline{C} \mid C \text{ connected component of } \widehat{\mathcal{S}}_I \smallsetminus \widehat{\mathcal{S}}_I^{\leq k},\ |I| = N,\ k \in \mathbb{N}\big\}.$$

Fix a finite binary word I. We will show that the following collection is null:

$$\mathcal{C}_I := \big\{\overline{C} \mid C \text{ connected component of } \widehat{\mathcal{S}}_I \smallsetminus \widehat{\mathcal{S}}_I^{\leq k},\ k \in \mathbb{N}\big\}.$$

This will complete the proof since $\mathcal{C}_N$ equals the finite union $\bigcup_{|I|=N} \mathcal{C}_I$.

First we show that for every nested strictly decreasing sequence of closures of components of complements of finite truncations of $\widehat{\mathcal{S}}_I$, that is a sequence $\overline{C}_1 \supsetneq \overline{C}_2 \supsetneq \cdots$ of elements of $\mathcal{C}_I$, the corresponding sequence of diameters converges to zero. For a contradiction, suppose that one such sequence $\{\operatorname{diam}\overline{C}_i\}$ does not converge to zero. So there exists $\varepsilon > 0$ such that, after passing to a subsequence, we have that $\operatorname{diam}\overline{C}_i \geq \varepsilon$ for each i. Choose points $a_i, b_i \in \overline{C}_i$ for each i with $d(a_i, b_i) \geq \varepsilon$. Since the ambient skyscraper $\widehat{\mathcal{S}}_I$ is a compact metric space, we assume, after passing to subsequences, that the sequences $\{a_i\}$ and $\{b_i\}$ converge. We call the limits a and b respectively, and observe that by construction they lie in $\bigcap \overline{C}_i$, and moreover $d(a,b) \geq \varepsilon$. This is a contradiction since $\bigcap \overline{C}_i$ is a single point, namely the endpoint of the ambient skyscraper $\widehat{\mathcal{S}}_I$ corresponding to the sequence $\{\overline{C}_i\}$.

Now we argue that the collection $\mathcal{C}_I$ is null, that is only finitely many elements have diameter $\geq \varepsilon$. For a contradiction, suppose there are infinitely many elements of $\mathcal{C}_I$ with diameter no less than ε. We will build a nested strictly decreasing sequence of closures of components of complements of finite truncations $\overline{C}_1 \supsetneq \overline{C}_2 \supsetneq \cdots$ of $\widehat{\mathcal{S}}_I$ such that $\operatorname{diam}\overline{C}_i \geq \varepsilon$ for all i, with $\overline{C}_i \in \mathcal{C}_I$. One of the (finitely many) components of $\widehat{\mathcal{S}}_I \smallsetminus \widehat{\mathcal{S}}_I^{\leq 1}$ must have infinitely many elements of $\mathcal{C}_I$ with diameter $\geq \varepsilon$, since otherwise there would be only finitely many such elements of $\mathcal{C}_I$. Assume we have chosen $\overline{C}_1, \dots, \overline{C}_j$ in this fashion, with each $\overline{C}_j$ a closure of a component of the complement $\widehat{\mathcal{S}}_I \smallsetminus \widehat{\mathcal{S}}_I^{\leq j}$, containing infinitely many elements of $\mathcal{C}_I$ with diameter $\geq \varepsilon$. Using the same argument as before, choose $\overline{C}_{j+1}$ a closure of a component of $\widehat{\mathcal{S}}_I \smallsetminus \widehat{\mathcal{S}}_I^{\leq j+1}$ such that $\overline{C}_{j+1} \subsetneq \overline{C}_j$ and so that $\overline{C}_j$ contains infinitely many elements of $\mathcal{C}_I$ with diameter $\geq \varepsilon$. By construction $\operatorname{diam}\overline{C}_i \geq \varepsilon$ for all i, as claimed, yielding the desired contradiction. This completes the proof that the collection $\mathcal{C}_I$, and thus that the collection of gaps$^+$ in $\widehat{\mathcal{S}}$, is null. $\square$

Let $\daleth^{-1}(\mathcal{G}^+)$ be the induced decomposition of D^4 obtained from the decomposition $\mathcal{G}^+$ of $\widehat{\mathcal{S}} \cup_{\mathrm{Id}} \mathcal{E}$. By the previous lemma, we deduce that $\daleth^{-1}(\mathcal{G}^+)$ is a null decomposition of D^4. More precisely, by Lemma 28.24 the decomposition $\daleth^{-1}(\mathcal{G}^+)$ is null with respect to the metric on D^4 induced from the metric on $\widehat{\mathcal{S}} \cup_{\mathrm{Id}} \mathcal{E}$ by $\daleth$. By Lemma 28.22, nullity is metric independent for compact spaces, so $\daleth^{-1}(\mathcal{G}^+)$ is also null with respect to the standard metric on D^4.

Now we will construct a map $f\colon D^4 \to D^4$ to which we shall apply the ball to ball theorem. Define ∂Q to be the (homeomorphic) image under α of $\partial H = S^3$, or equivalently under β of $\partial\widehat{\mathcal{S}}$ (see (B1-B2)). By the α shrink (Proposition 28.21) there is a homeomorphism $\overline{\alpha}\colon H \to Q$ that restricts to the map α on the boundary ∂H union the collar $C_-(\partial_- H \times [0, \frac{1}{2}])$. Using the product of the homeomorphisms $\alpha|_{\partial H}$ and $\beta|_{\partial\widehat{\mathcal{S}}}$ with the identity on $[0,1]$, we have the following diagram:

(28.9)
$$\begin{tikzcd}
\widehat{\mathcal{S}} \arrow[d, "\beta"'] & \underset{\substack{x\sim(x,0),\\ x\in\partial\widehat{\mathcal{S}}}}{\cup} & \partial\widehat{\mathcal{S}}\times[0,1] \arrow[d, "\beta\times\mathrm{Id}"] \\
Q & \underset{\substack{y\sim(y,0),\\ y\in\partial Q}}{\cup} & \partial Q\times[0,1] \\
H \arrow[u, "\alpha"] \arrow[u, bend left=70, "\overline{\alpha}"] & \underset{\substack{z\sim(z,0),\\ z\in\partial H}}{\cup} & S^3\times[0,1] \arrow[u, "\alpha\times\mathrm{Id}"']
\end{tikzcd}$$

The space $H \cup_{z\sim(z,0)} S^3 \times [0,1]$ is canonically homeomorphic to D^4. Let

$$\aleph\colon H \cup (S^3 \times [0,1]) \to Q \cup (\partial Q \times [0,1])$$

and

$$\widetilde{\beta}\colon \widehat{\mathcal{S}} \cup (\partial\widehat{\mathcal{S}} \times [0,1]) \to Q \cup (\partial Q \times [0,1])$$

denote the extensions of $\overline{\alpha}$ and β depicted in the diagram respectively. Then we define $f\colon D^4 \to D^4$ to be the composition

(28.10)

$$\begin{tikzcd} f\colon D^4 \arrow[r, "\daleth", "\cong"'] & \widehat{\mathcal{S}} \cup (\partial\widehat{\mathcal{S}} \times [0,1]) \arrow[dr, "\widetilde{\beta}"] & & H \cup (S^3 \times [0,1]) = D^4 \arrow[dl, "\aleph"', "\cong"] \\ & & Q \cup (\partial Q \times [0,1]) & \end{tikzcd}$$

We saw in Lemma 28.24 that the collection of inverse sets of f is null. To apply the ball to ball theorem we will also need that the collection of singular images of f is nowhere dense in the codomain.

Lemma 28.25. *The singular image of $\daleth^{-1}(\mathcal{G}^+)$ under $f\colon D^4 \to D^4$ is nowhere dense.*

Proof. Recall that a subset of a topological space is said to be nowhere dense if its closure has empty interior. First, since $\aleph$ is a homeomorphism and since $\widetilde{\beta}$ agrees with the homeomorphism $\beta \times \mathrm{Id}$ on $\mathcal{E} := \partial\widehat{\mathcal{S}} \times [0,1]$ (see (B2) and (28.9)), it suffices to show that the image of $\mathcal{G}^+$ in Q is nowhere dense. Next, since the image in Q of the gaps$^+$ in $\widehat{\mathcal{S}}$ is equal to the image in Q of the holes$^+$ in H, it suffices to show that the image in Q of the holes$^+$ in H is nowhere dense. This allows us to use the map α instead, and for this we can verify that the image of the holes$^+$ is nowhere dense in Q.

Now to complete the proof we argue that the condition in Proposition 28.23 is satisfied for the holes$^+$ in H. For the central hole$^+$, a neighbourhood V as in Proposition 28.23 is easily found within the collar of the attaching region $\partial_- H$.

Consider a hole$^+$ ${h_I^j}^+$ and let U be some open neighbourhood. We will now describe how to construct a neighbourhood V as in Proposition 28.23. By construction, ${h_I^j}^+$ equals the union of $L_I^j \times [r_{I\cdot 1}(0), r_{I\cdot 0}(1)]$ and a red blood cell disc $\widetilde{\Delta}_I^j$, where L_I^j is a tubular neighbourhood of a component of a mixed ramified Bing-Whitehead link $L_I \subseteq D^2 \times S^1$. Let T_I^j be the solid torus one stage up in the sequence of nested solid tori constructing L_I^j. In particular, $L_I^j \subsetneq T_I^j$ and forms a Whitehead double of the core of T_I^j.

We claim that only finitely many holes$^+$ may intersect the open set

$$W_I^j := \mathring{T}_I^j \times (r_{I\cdot 1}(0), r_{I\cdot 0}(1)).$$

To see this, observe that only the red blood cell discs for holes$^+$ corresponding to binary words J with length $|J| \leq |I|$ may intersect W_I^j (Lemma 28.19). This follows from the limits we put on how far red blood cell discs could be pushed during the pushing construction in Section 28.5: discs corresponding to longer indices cannot cross the r-coordinates $(r_{I\cdot 1}(0), r_{I\cdot 0}(1))$. Additionally, the holes intersecting W_I^j are precisely thickenings of the Whitehead doubles of parallel copies of the core of T_I^j, including h_I^j itself. This completes the proof of the claim.

Let V denote the intersection of U with W_I^j minus the finitely many holes$^+$ that intersect W_I^j. Since the holes$^+$ are closed, the nonempty set $V \subseteq U$ is open and does not meet any decomposition elements, including ${h_I^j}^+$. Apply Proposition 28.23

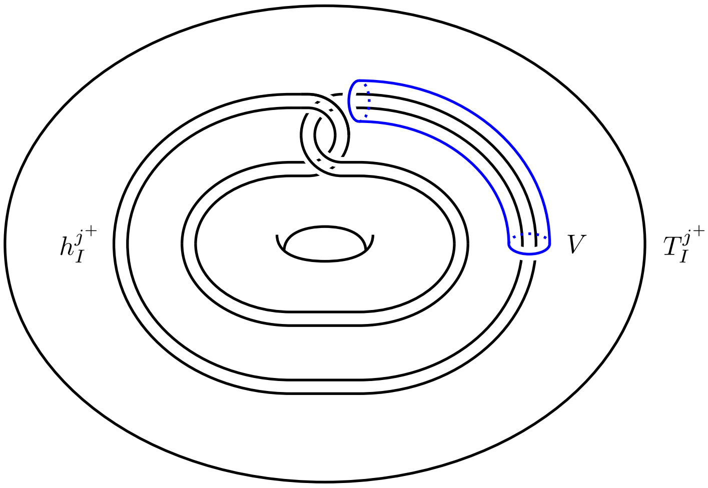


Figure 28.10. The proof that the singular image of f is nowhere dense.

using V to see that the image in Q of the holes$^+$ in H, and equivalently, the image in Q of the gaps$^+$ in $\widehat{\mathcal{S}}$, is nowhere dense. $\square$

Recall that $\mathcal{E} := \partial\widehat{\mathcal{S}} \times [0,1]$. Define an extension

$$\widetilde{\mathcal{E}} := \mathcal{E} \cup c_-(\partial_-\widehat{\mathcal{S}} \times [0, \tfrac{1}{2}]) \subseteq \widehat{\mathcal{S}}.$$

Define also

$$E := \aleph^{-1} \circ \widetilde{\beta}(\widetilde{\mathcal{E}}) \subseteq D^4.$$

Now we apply the ball to ball theorem (Theorem 26.5) to the map $f\colon D^4 \to D^4$ from (28.10) and the subset E. Let us check that the hypotheses hold. Recall that f is defined as the composition $\aleph^{-1} \circ \widetilde{\beta} \circ \gimel$. By construction we have $f^{-1}(E) = \gimel^{-1}(\widetilde{\mathcal{E}})$. Then $f|_{f^{-1}(E)}\colon f^{-1}(E) \to E$ is a homeomorphism, since both $\gimel$ and $\aleph$ are homeomorphisms and $\widetilde{\beta}$ restricts to a homeomorphism on $\widetilde{\mathcal{E}}$. For the latter assertion, see (28.9) and (B2). In addition, both E and $f^{-1}(E)$ contain $\partial D^4 = S^3$, so in particular $f|_{S^3}\colon S^3 \to S^3$ is a homeomorphism. As remarked in Chapter 10, this implies that f is degree one and so is surjective. By Lemma 28.24, the inverse sets $\gimel^{-1}(\mathcal{G}^+)$ are null in the domain D^4. By Lemma 28.25, the singular images are nowhere dense in the codomain D^4. Therefore all the hypotheses of the ball to ball theorem (Theorem 26.5) hold, so by the theorem the map f is approximable by homeomorphisms that agree with f on $f^{-1}(E) = \gimel^{-1}(\widetilde{\mathcal{E}})$.

Let $\overline{f}\colon D^4 \to D^4$ be such a homeomorphism approximating f. Then the composition

$$\widehat{\mathcal{S}} \cup_{\mathrm{Id}} \mathcal{E} \xrightarrow{\gimel^{-1}} D^4 \xrightarrow{\overline{f}} H \cup S^3 \times [0,1]$$

is a homeomorphism restricting to

$$f \circ \gimel^{-1} = \aleph^{-1} \circ \widetilde{\beta} \circ \gimel \circ \gimel^{-1} = \aleph^{-1} \circ \widetilde{\beta}$$

on $\widetilde{\mathcal{E}}$. Observe that the composition maps $\mathcal{E}$ to the added collar $S^3 \times [0,1]$ in the codomain (see (28.9)) by a homeomorphism, so by restricting to $\widehat{\mathcal{S}}$ we obtain a homeomorphism

$$\digamma\colon \widehat{\mathcal{S}} \to H = D^2 \times D^2.$$

We have shown that the map $f\colon D^4 \to D^4$ is approximable by homeomorphisms that agree with f on $f^{-1}(E)$. We also know that the maps $\aleph\colon D^4 \to Q \cup (\partial Q \times [0,1])$ and $\gimel^{-1}\colon \widehat{\mathcal{S}} \cup (\partial\widehat{\mathcal{S}} \times [0,1]) \to D^4$ are uniformly continuous by the Heine-Cantor theorem (Theorem 3.25). Thus, the map $\aleph \circ f \circ \gimel^{-1} = \aleph \circ \aleph^{-1} \circ \widetilde{\beta} \circ \gimel \circ \gimel^{-1} = \widetilde{\beta}$ is approximable by homeomorphisms agreeing with $\widetilde{\beta}$ on $\widetilde{\mathcal{E}}$, since $\gimel^{-1}(\widetilde{\mathcal{E}}) = f^{-1}(E)$. So, by restricting to $\widehat{\mathcal{S}}$ and recalling the definition of $\widetilde{\beta}$ from (28.9) we see that $\beta \colon \widehat{\mathcal{S}} \to Q$ is approximable by homeomorphisms agreeing with β on $\partial\widehat{\mathcal{S}} \cup c_-(\partial\widehat{\mathcal{S}} \times [0, \frac{1}{2}])$.

To finish the proof of Theorem 27.1, we need to understand the behaviour of the homeomorphism $\digamma$ close to the attaching region. We unravel some more definitions. Recall that the map $\aleph$ restricts to α on $C_-(\partial_- H \times [0, \frac{1}{2}])$ as defined in (28.9). Likewise the map $\widetilde{\beta}$ restricts to β on $c_-(\partial_-\widehat{\mathcal{S}} \times [0, \frac{1}{2}])$. So the restriction $\digamma|_{c_-(\partial_-\widehat{\mathcal{S}} \times [0, \frac{1}{2}])}$ agrees with the function $\alpha^{-1} \circ \beta|_{c_-\left(\partial_-\widehat{\mathcal{S}} \times [0, \frac{1}{2}]\right)}$, which is a diffeomorphism by (B3).

Finally, let $\partial\digamma := \digamma|_{\partial_-\widehat{\mathcal{S}}}$. Then if $\Phi \colon S^1 \times D^2$ is the attaching region of $\widehat{\mathcal{S}}$, we have that $\partial\digamma \circ \Phi = \alpha^{-1} \circ \beta \circ \Phi = \mathrm{Id}_{S^1 \times D^2}$ by (B4).

Summarising, we have constructed a homeomorphism of pairs

$$(\digamma, \partial\digamma) \colon \left(\widehat{\mathcal{S}}, \partial_-\widehat{\mathcal{S}}\right) \xrightarrow{\cong} \left(D^2 \times D^2, S^1 \times D^2\right)$$

that restricts to a diffeomorphism from a collar of $\partial_-\widehat{\mathcal{S}}$ to a collar of $S^1 \times D^2$, and such that for the attaching region $\Phi \colon S^1 \times D^2 \to \partial_-\widehat{\mathcal{S}}$ we have $\partial\digamma \circ \Phi = \mathrm{Id}_{S^1 \times D^2}$. This completes the proof of Theorem 27.1. Combining this with Section 18.4, in particular Proposition 18.12, completes the proof of the disc embedding theorem.

Bibliography

[Akb91] S. Akbulut, *A fake compact contractible* 4*-manifold*, J. Differential Geom. **33** (1991), no. 2, 335–356.

[Ale24] J. W. Alexander, *An example of a simply connected surface bounding a region which is not simply connected*, Proc. Natl. Acad. Sci. USA **10** (1924), 8–10.

[Ale30] J. W. Alexander, *The combinatorial theory of complexes*, Ann. of Math. (2) **31** (1930), no. 2, 292–320.

[Anc20] F. D. Ancel, *Recursively squeezable sets are squeezable*, 2020. Preprint, available at arXiv:2009.02817.

[Anc84] F. D. Ancel, *Approximating cell-like maps of S^4 by homeomorphisms*, Four-manifold theory (Durham, N.H., 1982), 1984, pp. 143–164.

[AP08] A. Akhmedov and B. D. Park, *Exotic smooth structures on small* 4*-manifolds*, Invent. Math. **173** (2008), no. 1, 209–223.

[AP10] A. Akhemedov and B. D. Park, *Exotic smooth structures on small* 4*-manifolds with odd signatures*, Invent. Math. **181** (2010), no. 3, 577–603.

[AR65] J. J. Andrews and L. Rubin, *Some spaces whose product with E^1 is E^4*, Bull. Amer. Math. Soc. **71** (1965), 675–677.

[Arm70] M. A. Armstrong, *Collars and concordances of topological manifolds*, Comment. Math. Helv. **45** (1970), 119–128.

[AS89] F. D. Ancel and M. P. Starbird, *The shrinkability of Bing-Whitehead decompositions*, Topology **28** (1989), no. 3, 291–304.

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

[BDK07] J. Brookman, J. F. Davis, and Q. Khan, *Manifolds homotopy equivalent to $P^n \# P^n$*, Math. Ann. **338** (2007), no. 4, 947–962.

[Bea67] R. J. Bean, *Decompositions of E^3 with a null sequence of starlike equivalent non-degenerate elements are E^3*, Illinois J. Math. **11** (1967), 21–23.

[BG64] M. Brown and H. Gluck, *Stable structures on manifolds. II. Stable manifolds*, Ann. of Math. (2) **79** (1964), 18–44.

[BG96] Ž. Bižaca and R. E. Gompf, *Elliptic surfaces and some simple exotic $\mathbf{R}^4$'s*, J. Differential Geom. **43** (1996), no. 3, 458–504.

[BHS64] H. Bass, A. Heller, and R. G. Swan, *The Whitehead group of a polynomial extension*, Inst. Hautes Études Sci. Publ. Math. **22** (1964), 61–79.

[Biž94] Ž. Bižaca, *A reimbedding algorithm for Casson handles*, Trans. Amer. Math. Soc. **345** (1994), no. 2, 435–510.

[Bin52] R. H. Bing, *A homeomorphism between the* 3*-sphere and the sum of two solid horned spheres*, Ann. of Math. (2) **56** (1952), 354–362.

[Bin59] R. H. Bing, *An alternative proof that* 3*-manifolds can be triangulated*, Ann. of Math. (2) **69** (1959), 37–65.

[Bin61] R. H. Bing, *Point-like decompositions of E^3*, Fund. Math. **50** (1961/1962), 431–453.

[Bin64] R. H. Bing, *Retractions onto spheres*, Amer. Math. Monthly **71** (1964), 481–484.

[Bin88] R. H. Bing, *Shrinking without lengthening*, Topology **27** (1988), no. 4, 487–493.

[BK08] S. Baldridge and P. Kirk, *A symplectic manifold homeomorphic but not diffeomorphic to $\mathbb{CP}^2 \# 3\overline{\mathbb{CP}^2}$*, Geom. Topol. **12** (2008), no. 2, 919–940.

[BL78] J. L. Bryant and R. C. Lacher, *Resolving zero-dimensional singularities in generalized manifolds*, Math. Proc. Cambridge Philos. Soc. **83** (1978), no. 3, 403–413.

[Boy86] S. Boyer, *Simply-connected* 4*-manifolds with a given boundary*, Trans. Amer. Math. Soc. **298** (1986), no. 1, 331–357.

[Boy93] S. Boyer, *Realization of simply-connected* 4*-manifolds with a given boundary*, Comment. Math. Helv. **68** (1993), no. 1, 20–47.

[Bro60] M. Brown, *A proof of the generalized Schoenflies theorem*, Bull. Amer. Math. Soc. **66** (1960), 74–76.

[Bro62] M. Brown, *Locally flat imbeddings of topological manifolds*, Ann. of Math. (2) **75** (1962), 331–341.

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

[BS70] J. L. Bryant and C. L. Seebeck III, *Locally nice embeddings in codimension three*, Quart. J. Math. Oxford Ser. (2) **21** (1970), 265–272.

[BV68] J. M. Boardman and R. M. Vogt, *Homotopy-everything H-spaces*, Bull. Amer. Math. Soc. **74** (1968), 1117–1122.

[BV73] J. M. Boardman and R. M. Vogt, *Homotopy invariant algebraic structures on topological spaces*, Lecture Notes in Mathematics, Vol. 347, Springer-Verlag, Berlin-New York, 1973.

[BZH14] G. Burde, H. Zieschang, and M. Heusener, *Knots*, extended edition, De Gruyter Studies in Mathematics, vol. 5, De Gruyter, Berlin, 2014.

[Cap71] S. Cappell, *A splitting theorem for manifolds and surgery groups*, Bull. Amer. Math. Soc. **77** (1971), 281–286.

[Car13] C. Carathéodory, *Zur Ränderzuordnung bei konformer Abbildung*, Göttingen Nachrichten (1913), 509–518.

[Cas86] A. Casson, *Three lectures on new infinite constructions in 4-dimensional manifolds*, à la recherche de la topologie perdue, 1986, pp. 201–244. With an appendix by L. Siebenmann.

[Cer70] J. Cerf, *La stratification naturelle des espaces de fonctions différentiables réelles et le théorème de la pseudo-isotopie*, Inst. Hautes Études Sci. Publ. Math. **39** (1970), 5–173.

[CF84] A. Casson and M. Freedman, *Atomic surgery problems*, Four-manifold theory (Durham, N.H., 1982), 1984, pp. 181–199.

[CFT09] T. D. Cochran, S. Friedl, and P. Teichner, *New constructions of slice links*, Comment. Math. Helv. **84** (2009), no. 3, 617–638.

[CG78] A. Casson and C. Gordon, *On slice knots in dimension three*, Algebraic and geometric topology (Proc. Sympos. Pure Math., Stanford Univ., Stanford, Calif., 1976), part 2, 1978, pp. 39–53.

[CG88] T. D. Cochran and R. E. Gompf, *Applications of Donaldson's theorems to classical knot concordance, homology 3-spheres and Property P*, Topology **27** (1988), no. 4, 495–512.

[CH90] T. D. Cochran and N. Habegger, *On the homotopy theory of simply connected four manifolds*, Topology **29** (1990), no. 4, 419–440.

[Cha14] J. C. Cha, *Amenable L^2-theoretic methods and knot concordance*, Int. Math. Res. Not. IMRN **17** (2014), 4768–4803.

[Cha65] L. Charlap, *Compact flat Riemannian manifolds. I*, Ann. of Math. (2) **81** (1965), 15–30.

[Cha74] T. A. Chapman, *Topological invariance of Whitehead torsion*, Amer. J. Math. **96** (1974), 488–497.

[Cha79] T. A. Chapman, *Locally homotopically unknotted embeddings of manifolds in codimension two are locally flat*, Topology **18** (1979), no. 4, 339–348.

[CHH13] T. D. Cochran, S. Harvey, and P. Horn, *Filtering smooth concordance classes of topologically slice knots*, Geom. Topol. **17** (2013), no. 4, 2103–2162.

[CHL09] T. D. Cochran, S. Harvey, and C. Leidy, *Knot concordance and higher-order Blanchfield duality*, Geom. Topol. **13** (2009), no. 3, 1419–1482.

[CHL11] T. D. Cochran, S. Harvey, and C. Leidy, *2-torsion in the n-solvable filtration of the knot concordance group*, Proc. Lond. Math. Soc. (3) **102** (2011), no. 2, 257–290.

[Cho80] C. Chou, *Elementary amenable groups*, Illinois J. Math. **24** (1980), no. 3, 396–407.

[CK17] J. C. Cha and M. H. Kim, *The bipolar filtration of topologically slice knots*, 2017. Preprint, available at arXiv:1710.07803.

[CKP20] J. C. Cha, M. H. Kim, and M. Powell, *A family of freely slice good boundary links*, Math. Ann. **376** (2020), no. 3-4, 1009–1030.

[CM19] D. Crowley and T. Macko, *On the cardinality of the manifold set*, Geom. Dedicata **200** (2019), 265–285.

[Coc90] T. D. Cochran, *Derivatives of links: Milnor's concordance invariants and Massey's products*, Mem. Amer. Math. Soc. **84** (1990), no. 427, x+73.

[COT03] T. D. Cochran, K. E. Orr, and P. Teichner, *Knot concordance, Whitney towers and L^2-signatures*, Ann. of Math. (2) **157** (2003), no. 2, 433–519.

[COT04] T. D. Cochran, K. E. Orr, and P. Teichner, *Structure in the classical knot concordance group*, Comment. Math. Helv. **79** (2004), no. 1, 105–123.

[CP16] J. C. Cha and M. Powell, *Casson towers and slice links*, Invent. Math. **205** (2016), no. 2, 413–457.

[CP19] A. Conway and M. Powell, *Characterisation of homotopy ribbon discs*, 2019. Preprint, available at arXiv:1902.05321.

[CS80] S. E. Cappell and J. L. Shaneson, *Link cobordism*, Comment. Math. Helv. **55** (1980), no. 1, 20–49.

[CS85] S. E. Cappell and J. L. Shaneson, *On* 4*-dimensional s-cobordisms*, J. Differential Geom. **22** (1985), no. 1, 97–115.

[CST12] J. Conant, R. Schneiderman, and P. Teichner, *Universal quadratic forms and Whitney tower intersection invariants*, Proceedings of the Freedman Fest, 2012, pp. 35–60.

[CT07] T. D. Cochran and P. Teichner, *Knot concordance and von Neumann ρ-invariants*, Duke Math. J. **137** (2007), no. 2, 337–379.

[Dav05] J. F. Davis, *The Borel/Novikov conjectures and stable diffeomorphisms of* 4*-manifolds*, Geometry and topology of manifolds, 2005, pp. 63–76.

[Dav07] R. Daverman, *Decompositions of manifolds*, AMS Chelsea Publishing, Providence, RI, 2007. Reprint of the 1986 original.

[DMF92] S. De Michelis and M. H. Freedman, *Uncountably many exotic $\mathbf{R}^4$'s in standard* 4*-space*, J. Differential Geom. **35** (1992), no. 1, 219–254.

[Don83] S. K. Donaldson, *An application of gauge theory to four-dimensional topology*, J. Differential Geom. **18** (1983), no. 2, 279–315.

[Don86] S. K. Donaldson, *Connections, cohomology and the intersection forms of* 4*-manifolds*, J. Differential Geom. **24** (1986), no. 3, 275–341.

[Don87a] S. K. Donaldson, *Irrationality and the h-cobordism conjecture*, J. Differential Geom. **26** (1987), no. 1, 141–168.

[Don87b] S. K. Donaldson, *The orientation of Yang-Mills moduli spaces and* 4*-manifold topology*, J. Differential Geom. **26** (1987), no. 3, 397–428.

[DS83] R. Denman and M. Starbird, *Shrinking countable decompositions of S^3*, Trans. Amer. Math. Soc. **276** (1983), no. 2, 743–756.

[DV09] R. Daverman and G. Venema, *Embeddings in manifolds*, Graduate Studies in Mathematics, vol. 106, American Mathematical Society, Providence, RI, 2009.

[DV16] A. Donald and F. Vafaee, *A slicing obstruction from the $\frac{10}{8}$ theorem*, Proc. Amer. Math. Soc. **144** (2016), no. 12, 5397–5405.

[Edw63] C. H. Edwards Jr., *Open* 3*-manifolds which are simply connected at infinity*, Proc. Amer. Math. Soc. **14** (1963), 391–395.

[Edw80] R. Edwards, *The topology of manifolds and cell-like maps*, Proceedings of the International Congress of Mathematicians (Helsinki, 1978), 1980, pp. 111–127.

[Edw84] R. Edwards, *The solution of the* 4*-dimensional annulus conjecture (after Frank Quinn)*, Four-manifold theory (Durham, N.H., 1982), 1984, pp. 211–264.

[EK71] R. D. Edwards and R. C. Kirby, *Deformations of spaces of imbeddings*, Ann. Math. (2) **93** (1971), 63–88.

[End95] H. Endo, *Linear independence of topologically slice knots in the smooth cobordism group*, Topology Appl. **63** (1995), no. 3, 257–262.

[Fer79] S. Ferry, *Homotoping ε-maps to homeomorphisms*, Amer. J. Math. **101** (1979), no. 3, 567–582.

[Fis60] G. M. Fisher, *On the group of all homeomorphisms of a manifold*, Trans. Amer. Math. Soc. **97** (1960), 193–212.

[FK16] M. Freedman and V. Krushkal, *Engel relations in 4-manifold topology*, Forum Math. Sigma **4** (2016), e22, 57.

[FK17] M. Freedman and V. Krushkal, *A homotopy$^+$ solution to the A-B slice problem*, J. Knot Theory Ramifications **26** (2017), no. 2, 1740018, 18.

[FK20] M. Freedman and V. Krushkal, *Engel groups and universal surgery models*, J. Topol. **13** (2020), no. 3, 1302–1316.

[FK78] M. Freedman and R. Kirby, *A geometric proof of Rochlin's theorem*, Algebraic and geometric topology (Proc. Sympos. Pure Math., Stanford Univ., Stanford, Calif., 1976), Part 2, 1978, pp. 85–97.

[FM66] R. H. Fox and J. W. Milnor, *Singularities of* 2*-spheres in* 4*-space and cobordism of knots*, Osaka J. Math. **3** (1966), 257–267.

[FM88] R. Friedman and J. W. Morgan, *On the diffeomorphism types of certain algebraic surfaces. I*, J. Differential Geom. **27** (1988), no. 2, 297–369.

[FNOP19] S. Friedl, M. Nagel, P. Orson, and M. Powell, *A survey of the foundations of four-manifold theory in the topological category*, 2019. Preprint, available at arXiv:1910.07372.

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

[Fre31] H. Freudenthal, *Über die Enden topologischer Räume und Gruppen*, Math. Z. **33** (1931), 692–713.

[Fre42] H. Freudenthal, *Neuaufbau der Endentheorie*, Ann. of Math. (2) **43** (1942), 261–279.

[Fre79] M. H. Freedman, *A fake* $S^3 \times \mathbf{R}$, Ann. of Math. (2) **110** (1979), no. 1, 177–201.

[Fre82a] M. Freedman, *The topology of four-dimensional manifolds*, J. Differential Geom. **17** (1982), no. 3, 357–453.

[Fre82b] M. H. Freedman, *A surgery sequence in dimension four; the relations with knot concordance*, Invent. Math. **68** (1982), no. 2, 195–226.

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

[Fre85] M. H. Freedman, *A new technique for the link slice problem*, Invent. Math. **80** (1985), no. 3, 453–465.

[Fre86a] M. H. Freedman, *Are the Borromean rings A-B-slice?*, Topology Appl. **24** (1986), no. 1-3, 143–145. Special volume in honor of R. H. Bing (1914–1986).

[Fre86b] M. H. Freedman, *A geometric reformulation of* 4*-dimensional surgery*, Topology Appl. **24** (1986), no. 1-3, 133–141. Special volume in honor of R. H. Bing (1914–1986).

[Fre88] M. H. Freedman, $\mathrm{Whitehead}_3$ *is a "slice" link*, Invent. Math. **94** (1988), no. 1, 175–182.

[Fre93] M. Freedman, *Link compositions and the topological slice problem*, Topology **32** (1993), no. 1, 145–156.

[FS85] R. Fintushel and R. Stern, *Pseudofree orbifolds*, Ann. of Math. (2) **122** (1985), no. 2, 335–364.

[FS98] R. Fintushel and R. J. Stern, *Knots, links, and* 4*-manifolds*, Invent. Math. **134** (1998), no. 2, 363–400.

[FT86] M. H. Freedman and L. R. Taylor, *A universal smoothing of four-space*, J. Differential Geom. **24** (1986), no. 1, 69–78.

[FT95a] M. H. Freedman and P. Teichner, 4*-manifold topology.* I. *Subexponential groups*, Invent. Math. **122** (1995), no. 3, 509–529.

[FT95b] M. H. Freedman and P. Teichner, 4*-manifold topology.* II. *Dwyer's filtration and surgery kernels*, Invent. Math. **122** (1995), no. 3, 531–557.

[Fur01] M. Furuta, *Monopole equation and the* $\frac{11}{8}$*-conjecture*, Math. Res. Lett. **8** (2001), no. 3, 279–291.

[GA70] F. González-Acuña, *Dehn's construction on knots*, Bol. Soc. Mat Mexicana (2) **15** (1970), 58–79.

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

[Gom83] R. E. Gompf, *Three exotic* $\mathbf{R}^4$*'s and other anomalies*, J. Differential Geom. **18** (1983), no. 2, 317–328.

[Gom84] R. E. Gompf, *Infinite families of Casson handles and topological disks*, Topology **23** (1984), no. 4, 395–400.

[Gom85] R. E. Gompf, *An infinite set of exotic* $\mathbf{R}^4$*'s*, J. Differential Geom. **21** (1985), no. 2, 283–300.

[Gom86] R. E. Gompf, *Smooth concordance of topologically slice knots*, Topology **25** (1986), no. 3, 353–373.

[Gre69] F. P. Greenleaf, *Invariant means on topological groups and their applications*, Van Nostrand Mathematical Studies, No. 16, Van Nostrand Reinhold Co., New York-Toronto, Ont.-London, 1969.

[Gri85] R. I. Grigorchuk, *Degrees of growth of finitely generated groups and the theory of invariant means*, Math. USSR Izvestiya **25** (1985), no. 2, 259–300.

[Gro81] M. Gromov, *Groups of polynomial growth and expanding maps*, Inst. Hautes Études Sci. Publ. Math. **53** (1981), 53–73.

[GS65] S. Gitler and J. D. Stasheff, *The first exotic class of BF*, Topology **4** (1965), 257–266.

[GS84] R. Gompf and S. Singh, *On Freedman's reimbedding theorems*, Four-manifold theory (Durham, N.H., 1982), 1984, pp. 277–309.

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

[GT04] S. Garoufalidis and P. Teichner, *On knots with trivial Alexander polynomial*, J. Differential Geom. **67** (2004), no. 1, 167–193.

[Ham18] I. Hambleton, *Orientable* 4*-dimensional Poincaré complexes have reducible Spivak fibrations*, 2018. Preprint, available at arXiv:1808.07157.

[Han54] S. Hanai, *On closed mappings*, Proc. Japan Acad. **30** (1954), 285–288.

[Hat02] A. Hatcher, *Algebraic topology*, Cambridge University Press, Cambridge, 2002.

[Hau30] F. Hausdorff, *Erweiterung einer Homöomorphie*, Fundam. Math. **16** (1930), 353–360.

[Hem76] J. Hempel, 3-*manifolds*, AMS Chelsea Publishing Series, Princeton University Press, 1976.

[HH19] I. Hambleton and A. Hildum, *Topological* 4-*manifolds with right-angled Artin fundamental groups*, J. Topol. Anal. **11** (2019), no. 04, 777–821.

[HH67] W.-c. Hsiang and W.-y. Hsiang, *Differentiable actions of compact connected classical groups. I*, Amer. J. Math. **89** (1967), 705–786.

[Him65] C. J. Himmelberg, *On the product of quotient spaces*, Amer. Math. Monthly **72** (1965), 1103–1106.

[HK12] M. Hedden and P. Kirk, *Instantons, concordance, and Whitehead doubling*, J. Diff. Geom. **91** (2012), 281–319.

[HK88] I. Hambleton and M. Kreck, *On the classification of topological* 4-*manifolds with finite fundamental group*, Math. Ann. **280** (1988), no. 1, 85–104.

[HK93] I. Hambleton and M. Kreck, *Cancellation results for* 2-*complexes and* 4-*manifolds and some applications*, Two-dimensional homotopy and combinatorial group theory, 1993, pp. 281–308.

[HKL16] M. Hedden, S.-G. Kim, and C. Livingston, *Topologically slice knots of smooth concordance order two*, J. Differential Geom. **102** (2016), no. 3, 353–393.

[HKT09] I. Hambleton, M. Kreck, and P. Teichner, *Topological* 4-*manifolds with geometrically two-dimensional fundamental groups*, J. Topol. Anal. **1** (2009), no. 2, 123–151.

[HKT94] I. Hambleton, M. Kreck, and P. Teichner, *Nonorientable* 4-*manifolds with fundamental group of order* 2, Trans. Amer. Math. Soc. **344** (1994), no. 2, 649–665.

[HLSX18] M. J. Hopkins, J. Lin, X. D. Shi, and Z. Xu, *Intersection forms of Spin 4-manifolds and the Pin(2)-equivariant Mahowald invariant*, 2018. Preprint, available at arXiv:1812.04052.

[HM74] M. W. Hirsch and B. Mazur, *Smoothings of piecewise linear manifolds*, Princeton University Press, Princeton, N. J.; University of Tokyo Press, Tokyo, 1974.

[Hom14] J. Hom, *The knot Floer complex and the smooth concordance group*, Comment. Math. Helv. **89** (2014), no. 3, 537–570.

[Hsu87] F. Hsu, 4-*dimensional topological bordism*, Topology Appl. **26** (1987), no. 3, 281–285.

[HT00] I. Hambleton and L. R. Taylor, *A guide to the calculation of the surgery obstruction groups for finite groups*, Surveys on surgery theory, Vol. 1, 2000, pp. 225–274.

[HW16] J. Hom and Z. Wu, *Four-ball genus bounds and a refinement of the Ozsváth-Szabó tau invariant*, J. Symplectic Geom. **14** (2016), no. 1, 305–323.

[JK06] B. Jahren and S. Kwasik, *Manifolds homotopy equivalent to* $\mathbb{RP}^4\#\mathbb{RP}^4$, Math. Proc. Cambridge Philos. Soc. **140** (2006), no. 2, 245–252.

[Ker69] M. A. Kervaire, *Smooth homology spheres and their fundamental groups*, Trans. Amer. Math. Soc. **144** (1969), 67–72.

[Kir68] R. C. Kirby, *On the set of non-locally flat points of a submanifold of codimension one*, Ann. of Math. (2) **88** (1968), 281–290.

[Kir69] R. C. Kirby, *Stable homeomorphisms and the annulus conjecture*, Ann. of Math. (2) **89** (1969), 575–582.

[Kir89] R. C. Kirby, *The topology of* 4-*manifolds*, Lecture Notes in Mathematics, vol. 1374, Springer-Verlag, Berlin, 1989.

[Kis64] J. M. Kister, *Microbundles are fibre bundles*, Ann. of Math. (2) **80** (1964), 190–199.

[KL04] M. Kreck and W. Lück, *The Novikov conjecture. Geometry and algebra*, Oberwolfach Seminars 33, Birkhäuser Basel, 2004.

[KL08] B. Kleiner and J. Lott, *Notes on Perelman's papers*, Geom. Topol. **12** (2008), no. 5, 2587–2855.

[Kot89] D. Kotschick, *On manifolds homeomorphic to* $\mathbf{C}\mathrm{P}^2\#8\overline{\mathbf{C}\mathrm{P}}^2$, Invent. Math. **95** (1989), no. 3, 591–600.

[KP14] D. Kasprowski and M. Powell, *Shrinking of toroidal decomposition spaces*, Fund. Math. **227** (2014), no. 3, 271–296.

[KPT17] M. H. Kim, M. Powell, and P. Teichner, *The round handle problem*, 2017. Preprint, available at arXiv:1706.09571.

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

[Kre99] M. Kreck, *Surgery and duality*, Ann. of Math. (2) **149** (1999), no. 3, 707–754.

[KS77] R. C. Kirby and L. C. Siebenmann, *Foundational essays on topological manifolds, smoothings, and triangulations*, Princeton University Press, Princeton, N.J., 1977.

With notes by John Milnor and Michael Atiyah, Annals of Mathematics Studies, No. 88.

[KT01] R. C. Kirby and L. R. Taylor, *A survey of* 4*-manifolds through the eyes of surgery*, Surveys on surgery theory, Vol. 2, 2001, pp. 387–421.

[Kug84] K. Kuga, *Representing homology classes of* $S^2 \times S^2$, Topology **23** (1984), no. 2, 133–137.

[Lüc02] W. Lück, *A basic introduction to surgery theory*, Topology of high-dimensional manifolds, No. 1, 2 (Trieste, 2001), 2002, pp. 1–224.

[Lan17] M. Land, *Reducibility of low dimensional Poincaré duality spaces*, 2017. Preprint, available at arXiv:1711.08179.

[Las70a] R. Lashof, *The immersion approach to triangulation*, Topology of Manifolds (Proc. Inst., Univ. of Georgia, Athens, Ga., 1969), 1970, pp. 52–56.

[Las70b] R. Lashof, *The immersion approach to triangulation and smoothing. I. Lees' immersion theorem*, Proc. Advanced Study Inst. on Algebraic Topology (Aarhus, 1970), Vol. II, 1970, pp. 282–322.

[Las70c] R. Lashof, *The immersion approach to triangulation and smoothing. II. Triangulations and smoothings*, Proc. Advanced Study Inst. on Algebraic Topology (Aarhus, 1970), Vol. II, 1970, pp. 323–355.

[Las71] R. Lashof, *The immersion approach to triangulation and smoothing*, Algebraic topology (Proc. Sympos. Pure Math., Vol. XXII, Univ. Wisconsin, Madison, Wis., 1970), 1971, pp. 131–164.

[Lee11] J. M. Lee, *Introduction to topological manifolds*, Second edition, Graduate Texts in Mathematics, vol. 202, Springer, New York, 2011.

[Lev66] J. P. Levine, *Polynomial invariants of knots of codimension two*, Ann. of Math. (2) **84** (1966), 537–554.

[Lev69] J. P. Levine, *Knot cobordism groups in codimension two*, Comment. Math. Helv. **44** (1969), 229–244.

[Lev77] J. Levine, *Knot modules. I*, Trans. Amer. Math. Soc. **229** (1977), 1–50.

[LL16] L. Lewark and A. Lobb, *New quantum obstructions to sliceness*, Proc. Lond. Math. Soc. (3) **112** (2016), no. 1, 81–114.

[Lob09] A. Lobb, *A slice genus lower bound from* $\mathrm{sl}(n)$ *Khovanov-Rozansky homology*, Adv. Math. **222** (2009), no. 4, 1220–1276.

[LT84] R. Lashof and L. Taylor, *Smoothing theory and Freedman's work on four-manifolds*, Algebraic topology, Aarhus 1982 (Aarhus, 1982), 1984, pp. 271–292.

[Maz59] B. Mazur, *On embeddings of spheres*, Bull. Amer. Math. Soc. **65** (1959), 59–65.

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

[McA54] L. F. McAuley, *On the aposyndetic decomposition of continua*, ProQuest LLC, Ann Arbor, MI, 1954. Ph.D. thesis, The University of North Carolina at Chapel Hill.

[MH73] J. Milnor and D. Husemoller, *Symmetric bilinear forms*, Springer-Verlag, New York-Heidelberg, 1973. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 73.

[Mil57] J. Milnor, *Isotopy of links. Algebraic geometry and topology*, A symposium in honor of S. Lefschetz, 1957, pp. 280–306.

[Mil58] J. Milnor, *On simply connected* 4*-manifolds*, Symposium internacional de topología algebraica International symposi um on algebraic topology, 1958, pp. 122–128.

[Mil63] J. W. Milnor, *Morse theory*, Princeton University Press, Princeton, N.J., 1963. Based on lecture notes by M. Spivak and R. Wells. Annals of Mathematics Studies, No. 51.

[Mil65] J. W. Milnor, *Lectures on the h-cobordism theorem*, Princeton University Press, Princeton, N.J., 1965. Notes by L. Siebenmann and J. Sondow.

[Mil68] J. Milnor, *Growth of finitely generated solvable groups*, J. Differential Geometry **2** (1968), 447–449.

[Moi52a] E. E. Moise, *Affine structures in* 3*-manifolds. IV. Piecewise linear approximations of homeomorphisms*, Ann. of Math. (2) **55** (1952), 215–222.

[Moi52b] E. E. Moise, *Affine structures in* 3*-manifolds. V. The triangulation theorem and Hauptvermutung*, Ann. of Math. (2) **56** (1952), 96–114.

[Moi77] E. E. Moise, *Geometric topology in dimensions* 2 *and* 3, Springer-Verlag, New York-Heidelberg, 1977. Graduate Texts in Mathematics, Vol. 47.

[MOR19] J. Meier, P. Orson, and A. Ray, *Null, recursively starlike-equivalent decompositions shrink*, 2019. Preprint, available at arXiv:1909.06165.

[Mor60] M. Morse, *A reduction of the Schoenflies extension problem*, Bull. Amer. Math. Soc. **66** (1960), 113–115.

[MS78] T. Matumoto and L. Siebenmann, *The topological s-cobordism theorem fails in dimension* 4 *or* 5, Math. Proc. Cambridge Philos. Soc. **84** (1978), no. 1, 85–87.

[MT07] J. Morgan and G. Tian, *Ricci flow and the Poincaré conjecture*, Clay Mathematics Monographs, vol. 3, American Mathematical Society, Providence, RI; Clay Mathematics Institute, Cambridge, MA, 2007.

[Mun00] J. R. Munkres, *Topology*, Prentice Hall, Inc., Upper Saddle River, NJ, 2000. Second edition.

[MY11] J.-B. Meilhan and A. Yasuhara, *Whitehead double and Milnor invariants*, Osaka J. Math. **48** (2011), no. 2, 371–381.

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

[OS03] P. Ozsváth and Z. Szabó, *Knot Floer homology and the four-ball genus*, Geom. Topol. **7** (2003), 615–639.

[OSS17] P. Ozsváth, A. Stipsicz, and Z. Szabó, *Concordance homomorphisms from knot Floer homology*, Adv. Math. **315** (2017), 366–426.

[OT13] W. F. Osgood and E. H. Taylor, *Conformal transformations on the boundaries of their regions of definitions*, Trans. Amer. Math. Soc. **14** (1913), no. 2, 277–298.

[Par05] J. Park, *Simply connected symplectic 4-manifolds with* $b_2^+ = 1$ *and* $c_1^2 = 2$, Invent. Math. **159** (2005), no. 3, 657–667.

[Per02] G. Perelman, *The entropy formula for the Ricci flow and its geometric applications*, 2002. Preprint, available at arXiv:0211159.

[Per03a] G. Perelman, *Finite extinction time for the solutions to the Ricci flow on certain three-manifolds*, 2003. Preprint, available at arXiv:0307245.

[Per03b] G. Perelman, *Ricci flow with surgery on three-manifolds*, 2003. Preprint, available at arXiv:0303109.

[Pes90] G. Peschke, *The theory of ends*, Nieuw Arch. Wisk. (4) **8** (1990), no. 1, 1–12.

[PRT20] M. Powell, A. Ray, and P. Teichner, *The 4-dimensional disc embedding theorem and dual spheres*, 2020. Preprint, available at arXiv:2006.05209.

[PSS05] J. Park, A. I. Stipsicz, and Z. Szabó, *Exotic smooth structures on* $\mathbb{CP}^2 \# 5\overline{\mathbb{CP}^2}$, Math. Res. Lett. **12** (2005), no. 5-6, 701–712.

[Qui79] F. Quinn, *Ends of maps. I*, Ann. of Math. (2) **110** (1979), no. 2, 275–331.

[Qui82a] F. Quinn, *Ends of maps. II*, Invent. Math. **68** (1982), no. 3, 353–424.

[Qui82b] F. Quinn, *Ends of maps. III. Dimensions* 4 *and* 5, J. Differential Geom. **17** (1982), no. 3, 503–521.

[Qui86] F. Quinn, *Isotopy of* 4*-manifolds*, J. Differential Geom. **24** (1986), no. 3, 343–372.

[Qui88] F. Quinn, *Topological transversality holds in all dimensions*, Bull. Amer. Math. Soc. (N.S.) **18** (1988), no. 2, 145–148.

[Rad25] T. Radó, *Über den Begriff der Riemannschen Fläche*, Acta Litt. Sci. Szeged **2** (1925), no. 101–121.

[Ran02] A. Ranicki, *Algebraic and geometric surgery*, Oxford Mathematical Monographs, The Clarendon Press, Oxford University Press, Oxford, 2002. Oxford Science Publications.

[Ran73] A. A. Ranicki, *Algebraic L-theory. I. Foundations*, Proc. London Math. Soc. (3) **27** (1973), 101–125.

[Ran80] A. Ranicki, *The algebraic theory of surgery. I. Foundations*, Proc. London Math. Soc. (3) **40** (1980), no. 1, 87–192.

[Ran81] A. Ranicki, *Exact sequences in the algebraic theory of surgery*, Mathematical Notes, vol. 26, Princeton University Press, Princeton, N.J.; University of Tokyo Press, Tokyo, 1981.

[Ran92] A. Ranicki, *Lower K- and L-theory*, London Mathematical Society Lecture Note Series, vol. 178, Cambridge University Press, Cambridge, 1992.

[Ras10] J. A. Rasmussen, *Khovanov homology and the slice genus*, Invent. Math. **182** (2010), no. 2, 419–447.

[Ray15] A. Ray, *Casson towers and filtrations of the smooth knot concordance group*, Algebr. Geom. Topol. **15** (2015), no. 2, 1119–1159.

[Roc52] V. A. Rochlin, *New results in the theory of four-dimensional manifolds*, Doklady Akad. Nauk SSSR (N.S.) **84** (1952), 221–224.

[RS72] C. P. Rourke and B. J. Sanderson, *Introduction to piecewise-linear topology*, Springer-Verlag, New York-Heidelberg, 1972. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 69.

[RS97] D. Ruberman and R. J. Stern, *A fake smooth* $\mathbf{C}P^2\#\mathbf{R}P^4$, Math. Res. Lett. **4** (1997), no. 2-3, 375–378.

[Rud76] W. Rudin, *Principles of mathematical analysis*, Third edition, McGraw-Hill Book Co., New York-Auckland-Düsseldorf, 1976. International Series in Pure and Applied Mathematics.

[Sav02] N. Saveliev, *Invariants for homology* 3*-spheres*, Encyclopaedia of Mathematical Sciences, vol. 140, Springer-Verlag, Berlin, 2002. Low-Dimensional Topology, I.

[Sco05] A. Scorpan, *The wild world of 4-manifolds*, American Mathematical Society, Providence, RI, 2005.

[Ser70] J.-P. Serre, *Cours d'arithmétique*, Collection SUP: "Le Mathématicien", vol. 2, Presses Universitaires de France, Paris, 1970.

[Sie70] L. C. Siebenmann, *Infinite simple homotopy types*, Nederl. Akad. Wetensch. Proc. Ser. A 73 = Indag. Math. **32** (1970), 479–495.

[Sie80] L. Siebenmann, *Amorces de la chirurgie en dimension quatre*: *un* $S^3 \times \mathbf{R}$ *exotique* (*d'après Andrew H. Casson et Michael h. Freedman*), Séminaire Bourbaki (1978/79), 1980, pp. Exp. No. 536, pp. 183–207.

[Sie82] L. Siebenmann, *La conjecture de Poincaré topologique en dimension* 4 (*d'après Michael. H. Freedman*) **92** (1982), 219–248.

[Sie73] L. Siebenmann, *Approximating cellular maps by homeomorphisms*, Topology **11** (1973), 271–294.

[Sma62] S. Smale, *On the structure of manifolds*, Amer. J. Math. **84** (1962), 387–399.

[Smy66] N Smythe, *Boundary links*, Topology Seminar, Wisconsin, 1965, 1966, pp. 69–72.

[Spi67] M. Spivak, *Spaces satisfying Poincaré duality*, Topology **6** (1967), 77–101.

[SS05] A. I. Stipsicz and Z. Szabó, *An exotic smooth structure on* $\mathbb{CP}^2\#6\overline{\mathbb{CP}^2}$, Geom. Topol. **9** (2005), 813–832.

[Sta62] J. Stallings, *The piecewise-linear structure of Euclidean space*, Proc. Cambridge Philos. Soc. **58** (1962), 481–488.

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

[Sto56] A. H. Stone, *Metrizability of decomposition spaces*, Proc. Amer. Math. Soc. **7** (1956), 690–700.

[Sto93] R. Stong, *Simply-connected* 4*-manifolds with a given boundary*, Topology Appl. **52** (1993), no. 2, 161–167.

[Sto94] R. Stong, *Existence of* π_1*-negligible embeddings in* 4*-manifolds. A correction to Theorem 10.5 of Freedmann and Quinn*, Proc. Amer. Math. Soc. **120** (1994), no. 4, 1309–1314.

[Sul96] D. P. Sullivan, *Triangulating and smoothing homotopy equivalences and homeomorphisms. Geometric Topology Seminar Notes*, The Hauptvermutung book, 1996, pp. 69–103.

[Tau87] C. H. Taubes, *Gauge theory on asymptotically periodic* 4*-manifolds*, J. Differential Geom. **25** (1987), no. 3, 363–430.

[Tei97] P. Teichner, *On the star-construction for topological* 4*-manifolds*, Geometric topology (Athens, GA, 1993), 1997, pp. 300–312.

[Vog82] P. Vogel, *Simply connected* 4*-manifolds*, Algebraic topology, 1981, 1982, pp. 116–119. Lectures given at the Topology Seminar held at the Matematisk Institut, Aarhus University, Aarhus, 1981.

[Wal16] C. T. C. Wall, *Differential topology*, Cambridge Studies in Advanced Mathematics, vol. 156, Cambridge University Press, Cambridge, 2016.

[Wal64] C. T. C. Wall, *On simply-connected* 4*-manifolds*, J. London Math. Soc. **39** (1964), 141–149.

[Wal65a] C. T. C. Wall, *Finiteness conditions for* CW*-complexes*, Ann. of Math. (2) **81** (1965), 56–69.

[Wal65b] C. T. C. Wall, *Open 3-manifolds which are 1-connected at infinity*, Quart. J. Math. Oxford Ser. (2) **16** (1965), 263–268.

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

[Wan95] Z. Wang, *Classification of closed nonorientable* 4*-manifolds with infinite cyclic fundamental group*, Math. Res. Lett. **2** (1995), no. 3, 339–344.

[Wei94] C. A. Weibel, *An introduction to homological algebra*, Cambridge Studies in Advanced Mathematics, vol. 38, Cambridge University Press, Cambridge, 1994.

[Whi34] J. H. C. Whitehead, *Certain theorems about three-dimensional manifolds (I)*, The Quarterly Journal of Mathematics **1** (1934), 308–320.

[Whi35] J. H. C. Whitehead, *A certain open manifold whose group is unity*, The Quarterly Journal of Mathematics **1** (1935), 268–279.

[Whi49] J. H. C. Whitehead, *On simply connected,* 4*-dimensional polyhedra*, Comment. Math. Helv. **22** (1949), 48–92.

[Whi44] H. Whitney, *The self-intersections of a smooth n-manifold in $2n$-space*, Ann. of Math. (2) **45** (1944), 220–246.

[Wil49] R. L. Wilder, *Topology of Manifolds*, American Mathematical Society Colloquium Publications, vol. 32, American Mathematical Society, New York, N. Y., 1949.

[Wil70] S. Willard, *General topology*, Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1970.

[Wol68] J. A. Wolf, *Growth of finitely generated solvable groups and curvature of Riemannian manifolds*, J. Differential Geometry **2** (1968), 421–446.

[Wri89] D. Wright, *Bing-Whitehead Cantor sets*, Fund. Math. **182** (1989), 105–116.

Index